

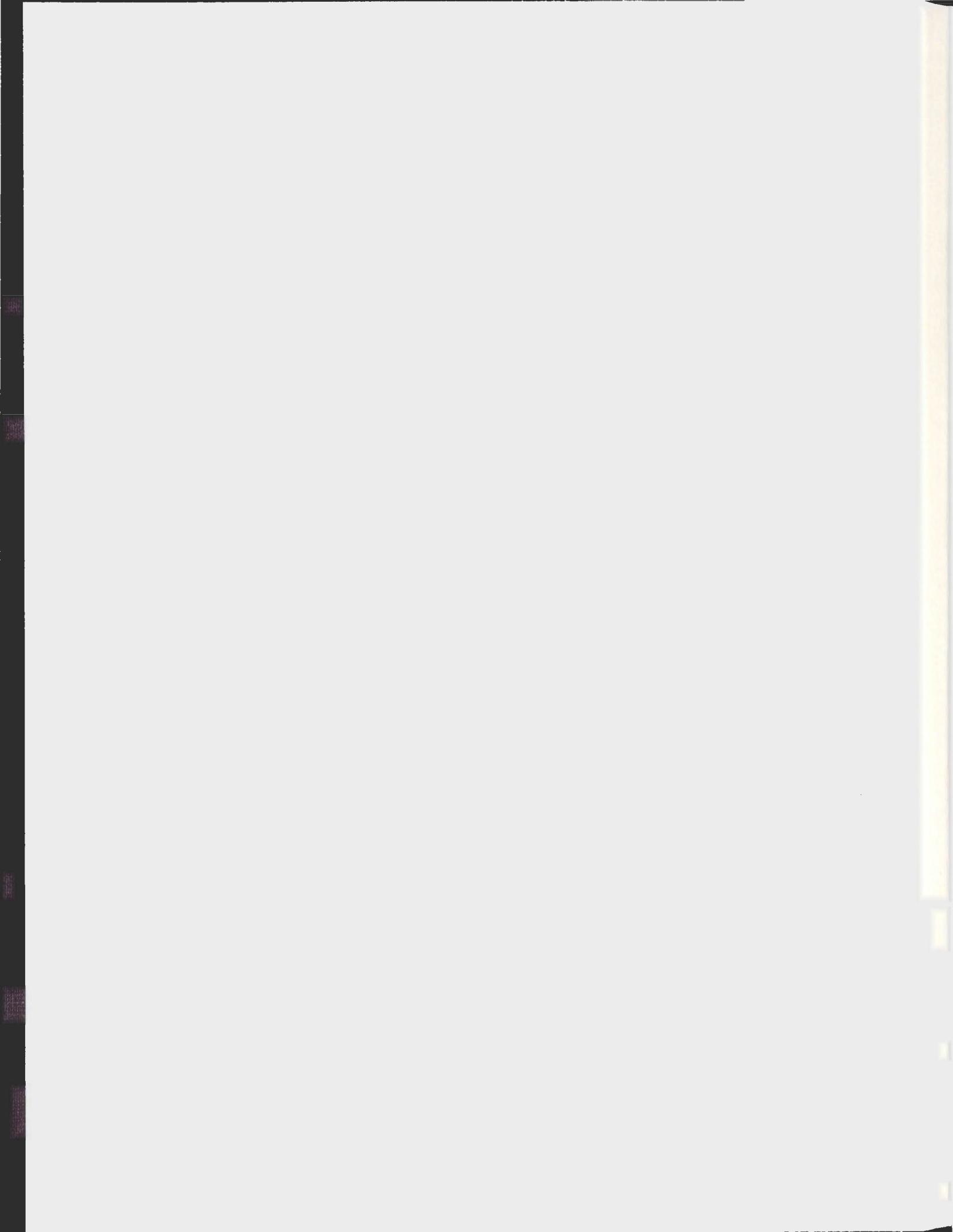
ON THE KUHN-TUCKER EQUIVALENCE THEOREM
AND ITS APPLICATIONS TO ISOTONIC REGRESSION

CENTRE FOR NEWFOUNDLAND STUDIES

**TOTAL OF 10 PAGES ONLY
MAY BE XEROXED**

(Without Author's Permission)

WEI LIU



INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

**ProQuest Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600**

UMI[®]



**National Library
of Canada**

**Acquisitions and
Bibliographic Services**

**385 Wellington Street
Ottawa ON K1A 0N4
Canada**

**Bibliothèque nationale
du Canada**

**Acquisitions et
services bibliographiques**

**385, rue Wellington
Ottawa ON K1A 0N4
Canada**

Your file Votre référence

Our file Notre référence

The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

L'auteur conserve la propriété du droit d'auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

0-612-66778-2

Canada

**On the Kuhn-Tucker Equivalence Theorem and its
Applications to Isotonic Regression**

by

Wei Liu

A thesis submitted to the
School of Graduate Studies
in partial fulfilment of the
requirements for the degree of
Master of Science

**Department of Mathematics and Statistics
Memorial University of Newfoundland
St. John's, Newfoundland**

July 2001

ABSTRACT

Prior information regarding a statistical model frequently constrains the shape of the parameter set and can often be quantified by placing inequality constraints on the parameters. For example, a regression function may be nondecreasing or convex or both; or the treatment response may stochastically dominate the control. The order restricted statistical inference has been well developed since the 1950's. The isotonic regression solves many restricted maximum likelihood estimation problems. And the theory of duality (cf. Barlow and Brunk (1972)) has provided insights into new problems. Both the isotonic regression and Fenchel duality play the important roles in order restricted statistical inference.

Kuhn and Tucker (1951) proposed a necessary and sufficient condition for the solution to an inequality constrained maximization problem. Since then, the Kuhn-Tucker equivalence theorem has been extensively applied to many fields such as optimization theory, engineering, the economy and so on.

In this paper, we focus on the applications of the Kuhn-Tucker equivalence theorem to order restricted estimation. This equivalence theorem provides a completely different approach to prove many important results such as the generalized isotonic regression problem due to Barlow and Brunk (1972), I-projection problems due to Dykstra (1985) and so on. We provide some insights into its extensive applications to ordered statistical inference. We expect that the kuhn-Tucker equivalence theorem

will become a powerful tool in this field.

Key words: Convex cone, Isotonic regression, Linearity space, Partial ordering, Polyhedral convex sets, Simple tree ordering, Van Eeden's algorithm.

ACKNOWLEDGMENTS

The author wishes to express her deepest appreciation to Professor Chu-In Charles Lee who initially suggested the problem and whose guidance has been very important throughout this research. The author considers herself very fortunate to have had the opportunity to study under his supervision. To Professor Lee the author expresses her wholehearted gratitude.

The author wishes to thank many members of the Department of Mathematics and Statistics with whom she has had valuable course work and discussion.

The author also acknowledges the financial assistance of the department under the headship of Dr. Herbert S. Gaskill and the School of Graduate Studies which allowed her to complete her educational program.

Finally thanks go to the author's husband, Enqing Shen, for his encouragement, understanding, patience and moral support.

TABLE OF CONTENTS

	Page
1 INTRODUCTION	1
2 PROPERTIES OF THE ISOTONIC REGRESSION	4
3 THE ISOTONIC CONES	7
4 THE GENERALIZED ISOTONIC REGRESSION PROBLEM	20
5 SOME RELATED PROBLEMS	31
5.1 EXPONENTIAL FAMILIES	31
5.2 A DYKSTRA-LEE THEOREM	33
5.3 I-PROJECTION PROBLEMS	35
5.4 MAXIMUM LIKELIHOOD ESTIMATIONS OF THE PROBABILITIES IN K BINOMIAL POPULATIONS	38
6 ON THE SIMPLE ORDERING	40
7 MULTINOMIAL ESTIMATION UNDER STOCHASTIC ORDERING	46
7.1 MAXIMUM LIKELIHOOD ESTIMATE ON MULTINOMIAL PARAMETERS	46
7.2 NEYMAN MODIFIED MINIMUM CHI-SQUARE ESTIMATE	53
REFERENCES	59

On the Kuhn-Tucker Equivalence Theorem and its Applications to Isotonic Regression

1. INTRODUCTION

Statistical inference is frequently involved in orderings and inequalities. For instance, the probability of a particular response may increase with the treatment level; the failure rate of a component may increase as it ages. The fact that the utilization of such ordering information increases the efficiency of statistical inference procedures is well documented. The one-tailed, two-sample t -test provides a familiar example in which the procedure which utilizes the prior information dominates procedures which ignore this information. Tests for identifying this structure often require good estimates under inequality constraints.

The isotonic regression solves many restricted maximum likelihood estimation problems (cf. Ayer et al. (1955), Van Eeden (1956, 1957a, 1957b, 1957c)). Perhaps the following (cf. Bartholomew (1959)) is the prototype problem leading to the isotonic regression problem. Given k normal distributions, $N(\mu_i, \sigma_i^2)$ ($i = 1, 2, \dots, k$), let \bar{x}_i be the sample mean from the i th distribution and n_i the sample size. One supposes that σ_i ($i = 1, 2, \dots, k$) are known and considers the problem of maximizing

the likelihood function subject to an order restriction on the means; i.e.,

$$\text{Maximize } \prod_{i=1}^k \frac{\sqrt{n_i}}{\sqrt{2\pi\sigma_i}} \exp[-n_i(\bar{x}_i - \mu_i)^2/2\sigma_i^2],$$

subject to $\mu_i \leq \mu_j$ when $i \prec j$,

where the binary relation \prec is a specified partial ordering on $\Omega = \{1, 2, \dots, k\}$. The partial ordering is provided by some prior knowledge concerning the means of the distributions.

This is equivalent to the problem

$$\text{Minimize } \sum_{i=1}^k (\bar{x}_i - \mu_i)^2 w_i, \tag{1.1}$$

subject to $\mu_i \leq \mu_j$ when $i \prec j$,

where $w_i = (n_i/\sigma_i^2) > 0$. This is called the isotonic regression problem since one seeks to minimize (1.1) subject to μ being an isotonic (i.e. order preserving) function on Ω . The solution to this problem is called the isotonic regression and is characterized in the next section. The solution to the isotonic regression problem is interesting because it also solves many other seemingly unrelated problems.

Barlow and Brunk (1972) formulated a generalization of the isotonic regression problem and studied its Fenchel duality. Their paper is an important milestone in the theory of order restricted inference. Since then, both the isotonic regression and Fenchel duality have become the critical tools in order restricted statistical inference.

Kuhn and Tucker (1951) provided a necessary and sufficient condition for the solution to an inequality constrained maximization problem. The Kuhn-Tucker equivalence theorem is one of the most important results in *mathematical programming*

and *operations research*. Applications to this equivalence theorem seem to have been centered largely in the econometrics literature where an active investigation has been taking place. Sources of work in this area are Gourieroux, Holly, and Monfort (1982), Farebrother (1986), and Wolak (1987). Although Kuhn and Tucker's result was published in 1951 and was extensively applied to many areas such as optimization theory, engineering, the economy and so on, their result has not been widely used in order restricted statistical inference.

Here we focus on the applications of the Kuhn-Tucker equivalence theorem to order restricted estimation. This equivalence theorem provides a completely different approach to proving many important results such as the generalized isotonic regression problem due to Barlow and Brunk (1972), I-projection problems due to Dykstra (1985) and so on. We hope that this important result will be more extensively applied in Statistics and become a powerful tool in order restricted statistical inference.

In this paper the properties of isotonic cones are discussed in Section 3. We solve the generalized isotonic regression problem due to Barlow and Brunk (1972) using the Kuhn-Tucker equivalence theorem in Section 4. Some related problems such as the maximum likelihood estimates for ordered parameters from an exponential family and I-projection with linear inequality constraints are solved in Section 5. Finally we consider some special cases in Section 6 and 7.

2. PROPERTIES OF THE ISOTONIC REGRESSION

The fact that the isotonic regression solves a wide class of restricted optimization problems was discovered independently by W. T. Reid and by Brunk, Ewing, and Utz (1957). Applications of their results to maximum likelihood estimation of ordered parameters were made in Brunk (1958). Further statistical applications were noted in Brunk (1965).

The isotonic regression problem is to

$$\text{Minimize } \sum_{i=1}^k (g_i - x_i)^2 w_i,$$

$$\text{subject to } x_i \leq x_j \text{ when } i < j,$$

where $w_i > 0$ and g_i ($i = 1, \dots, k$) are given and the binary relation $<$ is a specified partial ordering on $\Omega = \{1, 2, \dots, k\}$. We call the solution to this problem the *isotonic regression*. A vector $\mathbf{x} = (x_1, x_2, \dots, x_k)$ is said to be *isotonic* or *order preserving* on Ω with respect to the partial ordering $<$ if $i, j \in \Omega$ and $i < j$ implies $x_i \leq x_j$. Note that the set of isotonic functions $\mathbf{x} \in R^k$ is a closed convex cone, say K . This property is all that is needed for the following theorem.

Theorem 2.1. Let \mathbf{g} and \mathbf{w} be given functions on $\Omega = \{1, 2, \dots, k\}$, $\mathbf{w} > 0$. Let K be a closed convex cone of functions on Ω (not necessarily the cone of isotonic functions with respect to a partial ordering). Then

a. There is a unique $\mathbf{u} \in K$ solving

$$\text{Minimize } \sum_{i=1}^k (g_i - x_i)^2 w_i. \tag{2.1}$$

b. $\mathbf{u} \in K$ solves (2.1) if and only if

$$\sum_{i=1}^k (g_i - u_i)u_i w_i = 0, \quad (2.2)$$

and

$$\sum_{i=1}^k (g_i - u_i)x_i w_i \leq 0, \quad \forall \mathbf{x} \in K. \quad (2.3)$$

Notice that the above problem is equivalent to projecting \mathbf{g} onto K with respect to the inner product $\langle \mathbf{x}, \mathbf{y} \rangle_{\mathbf{w}} = \sum_{i=1}^k x_i y_i w_i$. The theorem follows immediately from Theorem 2.1 and Corollaries 2.3 and 2.1 of Brunk (1965).

We denote the solution to (2.1) by \mathbf{g}^* to indicate that it is the isotonic regression on \mathbf{g} with respect to the weights \mathbf{w} . If K is the cone of linear functions, for example, then \mathbf{g}^* becomes the ordinary linear regression.

The convex cone of isotonic functions is also a lattice. This is the additional property used to prove Theorem 2.2.

Theorem 2.2. Let \mathbf{g} and \mathbf{w} be given functions on $\Omega = \{1, 2, \dots, k\}$, $\mathbf{w} > 0$. Let K be a convex cone of isotonic functions on Ω with respect to some partial ordering \prec .

The isotonic regression \mathbf{g}^* of \mathbf{g} solving (2.1) exists and is unique.

If a_i and b_i are isotonic in i and $a_i \leq g_i \leq b_i, (i = 1, 2, \dots, k)$, then $a_i \leq g_i^* \leq b_i$.

A proof in a more general context may be found in Brunk (1965), and an elementary proof appeared in Barlow et al. (1972).

If K is a cone of isotonic functions, we can characterize \mathbf{g}^* in terms of upper and lower sets determined by the partial ordering \prec .

Definition 2.1. A subset L of Ω is a *lower* set with respect to \prec if $i \in L$ and $j \prec i$ implies $j \in L$. A subset U of Ω is an *upper* set if $i \in U$ and $i \prec j$ implies $j \in U$.

The solution to the isotonic regression problem can be characterized by

$$g_i^* = \max_{U: i \in U} \min_{L: i \in L} Av(L \cup U), \quad (2.4)$$

where for any nonempty subset $B \subseteq \Omega$,

$$Av(B) = \frac{\sum_{i \in B} w_i g_i}{\sum_{i \in B} w_i}.$$

This expression is given in Brunk et al. (1957) and in Brunk (1955).

3. THE ISOTONIC CONES

Now we introduce some definitions and notations for partial orderings. And then we will discuss the construction of isotonic cones.

For a partial ordering on Ω , if $i < j$ and there does not exist $r \in \Omega$ such that $i < r$ and $r < j$, then we express $i < j$ as $i <_I j$.

Set $E = \{(i, j); i <_I j\}$.

Note that the collection K of isotonic functions $\mathbf{x} \in R^k$ with respect to a partial ordering is a closed convex cone. We call K an isotonic cone.

Examples of isotonic cones include the simple ordering $(x_1 \leq x_2 \leq \dots \leq x_k)$, the simple tree ordering $(x_1 \leq x_j, \forall j)$, the unimodal ordering $(x_1 \leq x_2 \leq \dots \leq x_m \geq x_{m+1} \geq \dots \geq x_k)$, among others.

We can use graphs to express partial orderings. For example, $\Omega = \{1, 2, 3, 4\}$,

- simple ordering

$$1 < 2 < 3 < 4,$$

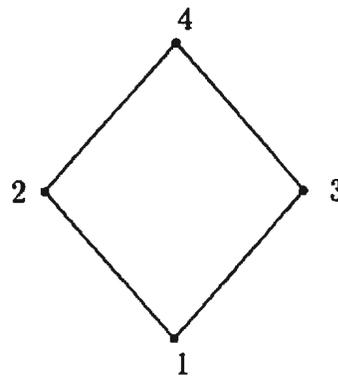
$$E = \{(1, 2), (2, 3), (3, 4)\}.$$



- simple loop

$$1 \prec [2, 3] \prec 4,$$

$$E = \{(1, 2), (1, 3), (2, 4), (3, 4)\}.$$



Definition 3.1. A partial ordering \prec is called connected if for every pair of u and v in Ω , there is a sequence $\{x_0, x_1, \dots, x_r\}$ such that $x_0 = u$, $x_r = v$ and either $x_i \prec x_{i+1}$ or $x_{i+1} \prec x_i$. Otherwise \prec is said to be disconnected.

Comparing the above two examples, both of them are connected. For a connected partial ordering, if $\#E = k - 1$, we call it a tree. If $\#E \geq k$, we say it contains cycles. For example, the simple ordering is a tree, but the simple loop contains one cycle.

For a partial ordering on Ω , K can be expressed as

$$\begin{aligned} K &= \{\mathbf{x} : x_i \leq x_j, (i, j) \in E\} \\ &= \{\mathbf{x} : x_j - x_i \geq 0, (i, j) \in E\} \\ &= \{\mathbf{x} : \langle \mathbf{a}_{ij}, \mathbf{x} \rangle \geq 0, (i, j) \in E\}, \end{aligned}$$

where $\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{i=1}^k x_i y_i$, $a_{ijl} = -1, 1, 0$ if $l = i, j$, otherwise (o/w), respectively.

Lemma 3.1. (Dual Basis). Let $\mathbf{a}_1, \dots, \mathbf{a}_n$ ($n \leq k$) be linearly independent and let $S (\subseteq R^k)$ be the linear subspace generated by $\mathbf{a}_1, \dots, \mathbf{a}_n$. Then there exist unique $\mathbf{b}_1, \dots, \mathbf{b}_n$ in S such that

$$\langle \mathbf{b}_i, \mathbf{a}_j \rangle = \begin{cases} 1, & i = j, \\ 0, & \text{o/w,} \end{cases} \quad (3.1)$$

and $\mathbf{b}_1, \dots, \mathbf{b}_n$ are linearly independent. $(\mathbf{b}_1, \dots, \mathbf{b}_n)$ is called the dual basis of $(\mathbf{a}_1, \dots, \mathbf{a}_n)$.

Proof. Since $\mathbf{a}_1, \dots, \mathbf{a}_n$ are linearly independent,

$$A = \begin{pmatrix} \mathbf{a}'_1 \\ \mathbf{a}'_2 \\ \vdots \\ \mathbf{a}'_n \end{pmatrix}$$

is a $n \times k$ matrix with full row rank. We expand the matrix A into the $k \times k$ non-singular matrix \bar{A} by adding the $(k - n)$ linearly independent vectors in S^\perp to rows of A .

For \mathbf{b}_1 , we have $\langle \mathbf{b}_1, \mathbf{a}_1 \rangle = 1$, $\langle \mathbf{b}_1, \mathbf{a}_i \rangle = 0$, $i = 2, \dots, n$. Then \mathbf{b}_1 satisfies

$$\bar{A} \mathbf{b}_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}_{k \times 1}.$$

So there exists a unique \mathbf{b}_1 satisfying the condition

$$\langle \mathbf{b}_1, \mathbf{a}_1 \rangle = 1, \quad \langle \mathbf{b}_1, \mathbf{a}_i \rangle = 0, \quad i = 2, \dots, n.$$

Similarly, there exist unique $\mathbf{b}_2, \dots, \mathbf{b}_n$ satisfying the equalities (3.1).

Consider a linear combination of $\mathbf{b}_1, \dots, \mathbf{b}_n$, $\lambda_1 \mathbf{b}_1 + \dots + \lambda_n \mathbf{b}_n$.

If $\lambda_1 \mathbf{b}_1 + \dots + \lambda_n \mathbf{b}_n = \mathbf{0}$, then

$$\langle \mathbf{a}_i, \lambda_1 \mathbf{b}_1 + \dots + \lambda_n \mathbf{b}_n \rangle = 0, \quad i = 1, \dots, n.$$

So

$$\lambda_i = 0, \quad i = 1, \dots, n.$$

Therefore, $\mathbf{b}_1, \dots, \mathbf{b}_n$ are linearly independent. □

Note that we separate the isotonic cones on the basis of whether they contain only trees or cycles. Now we will characterize the isotonic cones containing only trees.

Theorem 3.1. Let $\mathbf{a}_1, \dots, \mathbf{a}_n$ ($n \leq k$) be linearly independent and let $S (\subseteq R^k)$ be the linear subspace generated by $\mathbf{a}_1, \dots, \mathbf{a}_n$. Assume that $(\mathbf{b}_1, \dots, \mathbf{b}_n)$ is the dual basis of $(\mathbf{a}_1, \dots, \mathbf{a}_n)$. If $\mathbf{x} \in K = \{\mathbf{x} : \langle \mathbf{a}_i, \mathbf{x} \rangle \geq 0, i = 1, \dots, n\} \subseteq S$, then

$$\mathbf{x} = \sum_{i=1}^n \alpha_i \mathbf{b}_i, \quad \text{all } \alpha_i \geq 0.$$

Proof. Since $\mathbf{b}_1, \dots, \mathbf{b}_n$ are linearly independent by Lemma 3.1, for any $\mathbf{x} \in K (\subseteq S)$, there exist $\alpha_i \in R, i = 1, \dots, n$, such that

$$\mathbf{x} = \sum_{i=1}^n \alpha_i \mathbf{b}_i.$$

By the definition of the dual basis, we have

$$\begin{aligned} 0 \leq \langle \mathbf{a}_j, \mathbf{x} \rangle &= \langle \mathbf{a}_j, \sum_{i=1}^n \alpha_i \mathbf{b}_i \rangle \\ &= \sum_{i=1}^n \alpha_i \langle \mathbf{a}_j, \mathbf{b}_i \rangle = \alpha_j, \quad \forall j = 1, \dots, n. \end{aligned}$$

Hence $\alpha_j \geq 0, j = 1, \dots, n$. The result of this theorem follows. □

Theorem 3.2. Suppose a partial ordering corresponding to $E = \{(i, j); i \prec_I j\}$ contains only $m(\geq 1)$ disconnected trees and $\#E = q$. Then $q + m = k$. Also, q constraints \mathbf{a}_{ij} 's corresponding to E are linearly independent.

Proof. Since the partial ordering contains only trees, it is straightforward that $q+m = k$. And there exist p disconnected trees corresponding to nonempty sets E_α , $\alpha = 1, \dots, p$ ($\leq m$), such that

$$E = \bigcup_{\alpha=1}^p E_\alpha.$$

Let Ω_α be the subset of Ω corresponding to E_α , $\alpha = 1, \dots, p$. It is evident that $\Omega_\alpha \cap \Omega_\beta = \emptyset$, $\alpha \neq \beta$, $1 \leq \alpha, \beta \leq p$.

Let $\#E_\alpha = q_\alpha$. Then $\sum_{\alpha=1}^p q_\alpha = q$.

For any E_α ($1 \leq \alpha \leq p$), we can express E_α as

the first element $(r_1^{(\alpha)}, r_2^{(\alpha)})$ with $a_{r_1^{(\alpha)}, r_2^{(\alpha)}, l} = -1, 1, 0$ if $l = r_1^{(\alpha)}$, $l = r_2^{(\alpha)}$, o/w , respectively. And

the j th element $(x_j^{(\alpha)}, r_j^{(\alpha)})$ with $a_{x_j^{(\alpha)}, r_j^{(\alpha)}, l} = -1, 1, 0$ if $l = x_j^{(\alpha)}$, $l = r_j^{(\alpha)}$, o/w , respectively or $(r_j^{(\alpha)}, x_j^{(\alpha)})$ with $a_{r_j^{(\alpha)}, x_j^{(\alpha)}, l} = -1, 1, 0$ if $l = r_j^{(\alpha)}$, $l = x_j^{(\alpha)}$, o/w , respectively, where $x_j^{(\alpha)} \in \{r_1^{(\alpha)}, r_2^{(\alpha)}, \dots, r_{j-1}^{(\alpha)}\}$, $j = 2, \dots, q_\alpha$. Thus \mathbf{a}_{ij} 's, $(i, j) \in E_\alpha$, are linearly independent.

Since $\Omega_\alpha \cap \Omega_\beta = \emptyset$, $\alpha \neq \beta$, $1 \leq \alpha, \beta \leq p$, q constraints \mathbf{a}_{ij} 's corresponding to E are linearly independent. \square

Theorem 3.3. Suppose a partial ordering corresponding to E contains only m (≥ 1) disconnected trees and $\#E = q$. Then, we have

$$K = \left\{ \mathbf{x} : \mathbf{x} = \sum_{\alpha=1}^m \gamma_\alpha \mathbf{c}_\alpha + \sum_{(i,j) \in E} \nu_{ij} \mathbf{b}_{ij}, \quad \text{all } \nu_{ij} \geq 0 \right\}, \quad (3.2)$$

where \mathbf{c}_α 's are k -dimensional vectors defined as

$$\mathbf{c}_{\alpha r} = \begin{cases} 1, & r \in \Omega_\alpha, \\ 0, & o/w, \end{cases} \quad \alpha = 1, \dots, m,$$

with Ω_α being the subsets of Ω corresponding to disconnected trees, and $\{\mathbf{b}_{ij}, (i, j) \in E\}$ is the dual basis of $\{\mathbf{a}_{ij}, (i, j) \in E\}$.

Proof. K is a polyhedral convex cone and contains lines. Thus, K can be expressed as

$$K = L + K_0,$$

where L is the linearity space of K and K_0 is a closed convex set containing no lines, namely $K_0 = K \cap L^\perp$ (cf. Rockafellar (1970), the proof of Theorem 19.1, p171).

And

$$\begin{aligned} L = (-K) \cap K &= \{\mathbf{x} : \langle \mathbf{a}_{ij}, \mathbf{x} \rangle = 0, (i, j) \in E\} \\ &= \{\mathbf{x} : x_r = x_s, r, s \in \Omega_\alpha, \alpha = 1, \dots, m\}. \end{aligned}$$

It is evident that m \mathbf{c}_i 's are linearly independent. We will show that

$$L = \{\mathbf{x} : \mathbf{x} = \sum_{\alpha=1}^m \gamma_\alpha \mathbf{c}_\alpha\}. \quad (3.3)$$

In fact, apparently, the right hand side of (3.3) $\subseteq L$.

On the other hand, for $\forall \mathbf{x} \in L$, we can write \mathbf{x} as

$$\mathbf{x} = \sum_{\alpha=1}^m x_{i \in \Omega_\alpha} \mathbf{c}_\alpha$$

which belongs to the right hand side of (3.3). And $\dim(L) = m = k - q$ by Theorem 3.2.

Next we consider $K_0 = K \cap L^\perp$. Of course, $\mathbf{a}_{ij} \in L^\perp$, $(i, j) \in E$. And by Theorem 3.2, \mathbf{a}_{ij} 's are linearly independent.

Therefore, by Theorem 3.1,

$$K_0 = \{\mathbf{x} : \mathbf{x} = \sum_{(i,j) \in E} \nu_{ij} \mathbf{b}_{ij}, \text{ all } \nu_{ij} \geq 0\}.$$

Finally the result of this theorem follows. \square

For a partial ordering containing only trees, we denote U_{ij} and L_{ij} as the largest connected subsets of Ω containing j but not i and containing i but not j , respectively, $(i, j) \in E$. And let $I_{U_{ij}}$ and $I_{L_{ij}}$ be the indicators of U_{ij} and L_{ij} , respectively. Then we will have the following theorem.

Theorem 3.4. In expression (3.2),

$$\mathbf{b}_{ij} = \frac{\#L_{ij}}{\#L_{ij} + \#U_{ij}} I_{U_{ij}} - \frac{\#U_{ij}}{\#L_{ij} + \#U_{ij}} I_{L_{ij}}, \quad (3.4)$$

$$\nu_{ij} = x_j - x_i, \quad (i, j) \in E.$$

Proof. Note that \mathbf{b}_{ij} , $(i, j) \in E$, in Theorem 3.3 satisfy the following conditions

$$\langle \mathbf{b}_{ij}, \mathbf{a}_{i'j'} \rangle = \begin{cases} 1, & i = i', j = j', \\ 0, & \text{o/w,} \end{cases} \quad (i, j), (i', j') \in E,$$

$$\langle \mathbf{b}_{ij}, \mathbf{x} \rangle = 0, \quad \forall \mathbf{x} \in L.$$

On the construction of \mathbf{a}_{ij} , $I_{U_{ij}}$ and $I_{L_{ij}}$, $(i, j) \in E$, it is evident that the right hand side of (3.4) satisfies the above conditions.

Since there exists the unique dual basis of $\{\mathbf{a}_{ij}, (i, j) \in E\}$ by Lemma 3.1,

$$\mathbf{b}_{ij} = \frac{\#L_{ij}}{\#L_{ij} + \#U_{ij}} I_{U_{ij}} - \frac{\#U_{ij}}{\#L_{ij} + \#U_{ij}} I_{L_{ij}}, \quad (i, j) \in E.$$

For $\forall \mathbf{x} \in K$, by Theorem 3.3 we have

$$\mathbf{x} = \sum_{\alpha=1}^m \gamma_{\alpha} \mathbf{c}_{\alpha} + \sum_{(i', j') \in E} \nu_{i'j'} \mathbf{b}_{i'j'}, \quad \text{all } \nu_{i'j'} \geq 0.$$

For $\forall \mathbf{a}_{ij}, (i, j) \in E$,

$$\langle \mathbf{a}_{ij}, \mathbf{x} \rangle = x_j - x_i.$$

On the other hand,

$$\begin{aligned} & \langle \mathbf{a}_{ij}, \sum_{\alpha=1}^m \gamma_{\alpha} \mathbf{c}_{\alpha} + \sum_{(i', j') \in E} \nu_{i'j'} \mathbf{b}_{i'j'} \rangle \\ &= \sum_{\alpha=1}^m \gamma_{\alpha} \langle \mathbf{a}_{ij}, \mathbf{c}_{\alpha} \rangle + \sum_{(i', j') \in E} \nu_{i'j'} \langle \mathbf{a}_{ij}, \mathbf{b}_{i'j'} \rangle \\ &= \nu_{ij}. \end{aligned}$$

Therefore, $\nu_{ij} = x_j - x_i$. □

By Theorem 3.4, we can display K as

$$K = \left\{ \mathbf{x} : \mathbf{x} = \sum_{\alpha=1}^m \gamma_{\alpha} \mathbf{c}_{\alpha} + \sum_{(i, j) \in E} (x_j - x_i) \left(\frac{\#L_{ij}}{\#L_{ij} + \#U_{ij}} I_{U_{ij}} - \frac{\#U_{ij}}{\#L_{ij} + \#U_{ij}} I_{L_{ij}} \right) \right\}. \quad (3.5)$$

Example 3.1. Consider $\Omega = \{1, 2, 3, 4\}$ and the simple tree: $1 \prec [2, 3, 4]$. We can write $E = \{(1, 2), (1, 3), (1, 4)\}$. The simple tree cone is expressed as

$$\begin{aligned} K &= \{\mathbf{x} : x_1 \leq x_i, \quad i = 2, 3, 4\} \\ &= \{\mathbf{x} : \langle \mathbf{a}_{ij}, \mathbf{x} \rangle \geq 0, \quad (i, j) \in E\}, \end{aligned}$$

where the constraints

$$\mathbf{a}_{12} = (-1, 1, 0, 0)$$

$$\mathbf{a}_{13} = (-1, 0, 1, 0)$$

$$\mathbf{a}_{14} = (-1, 0, 0, 1).$$

And the corresponding U_{ij} , $I_{U_{ij}}$ and $\#U_{ij}$ are

$$U_{12} = \{2\}, \quad I_{U_{12}} = (0, 1, 0, 0), \quad \#U_{12} = 1$$

$$U_{13} = \{3\}, \quad I_{U_{13}} = (0, 0, 1, 0), \quad \#U_{13} = 1$$

$$U_{14} = \{4\}, \quad I_{U_{14}} = (0, 0, 0, 1), \quad \#U_{14} = 1.$$

Similarly, the corresponding L_{ij} , $I_{L_{ij}}$ and $\#L_{ij}$ are

$$L_{12} = \{1, 3, 4\}, \quad I_{L_{12}} = (1, 0, 1, 1), \quad \#L_{12} = 3$$

$$L_{13} = \{1, 2, 4\}, \quad I_{L_{13}} = (1, 1, 0, 1), \quad \#L_{13} = 3$$

$$L_{14} = \{1, 2, 3\}, \quad I_{L_{14}} = (1, 1, 1, 0), \quad \#L_{14} = 3.$$

Therefore, we can obtain the dual basis of \mathbf{a}_{ij} 's as

$$\mathbf{b}_{12} = \frac{1}{4}(-1, 3, -1, -1),$$

$$\mathbf{b}_{13} = \frac{1}{4}(-1, -1, 3, -1),$$

$$\mathbf{b}_{14} = \frac{1}{4}(-1, -1, -1, 3).$$

The only one $\mathbf{c} = \mathbf{1}_4 = (1, 1, 1, 1)'$. We can express K in the form of (3.5) as

$$K = \{\mathbf{x} : \mathbf{x} = \bar{x} \mathbf{1}_4 + (x_2 - x_1) \mathbf{b}_{12} + (x_3 - x_1) \mathbf{b}_{13} + (x_4 - x_1) \mathbf{b}_{14}\},$$

where \bar{x} is the mean of \mathbf{x} .

We have observed the construction of isotonic cones containing only trees. This will be used in the following section. Van Eeden (1958) provided the following important result which supports Van Eeden's algorithm. Although her algorithm is not feasible for a more complicated partial ordering, her result is valuable in theory.

Lemma 3.2. (Van Eeden). Suppose that a partial ordering on Ω corresponds to E and $\#E = q$. Order $E = \{(i_1, j_1), (i_2, j_2), \dots, (i_q, j_q)\}$ and let $\bar{E} = \{(i_1, j_1), (i_2, j_2), \dots, (i_{q-1}, j_{q-1})\}$. Then for any given function \mathbf{x} on Ω , there are two possible cases.

(a) $[E_{\mathbf{w}}(\mathbf{x}|K_E)]_{i_q} \leq [E_{\mathbf{w}}(\mathbf{x}|K_E)]_{j_q}$, then

$$E_{\mathbf{w}}(\mathbf{x}|K_E) = E_{\mathbf{w}}(\mathbf{x}|K_{\bar{E}});$$

(b) $[E_{\mathbf{w}}(\mathbf{x}|K_E)]_{i_q} > [E_{\mathbf{w}}(\mathbf{x}|K_E)]_{j_q}$, then

$$[E_{\mathbf{w}}(\mathbf{x}|K_E)]_{i_q} = [E_{\mathbf{w}}(\mathbf{x}|K_E)]_{j_q},$$

where K_E is the isotonic cone corresponding to E , and $E_{\mathbf{w}}(\mathbf{x}|K)$ is the isotonic regression of \mathbf{x} with respect to the weights \mathbf{w} .

We mainly discussed the construction of isotonic cones containing only trees. For partial orderings containing cycles, we have the following theorem.

Theorem 3.5. If a partial ordering corresponding to E_c contains cycles, then for any given function \mathbf{x} on Ω , there exists a consistent partial ordering corresponding to $E_t \subset E_c$, which contains only trees, such that

$$E_{\mathbf{w}}(\mathbf{x}|K_{E_c}) = E_{\mathbf{w}}(\mathbf{x}|K_{E_t}),$$

where K_E is the isotonic cone corresponding to E , and $E_{\mathbf{w}}(\mathbf{x}|K)$ is the isotonic regression of \mathbf{x} with respect to the weights \mathbf{w} .

Proof. Let $\#E_c = q$. First we order $E_c = \{(i, j); i \prec_t j\} = \{(i_1, j_1), (i_2, j_2), \dots, (i_q, j_q)\}$

and let $\bar{E}_c = \{(i_1, j_1), (i_2, j_2), \dots, (i_{q-1}, j_{q-1})\}$. We will proceed with the following iteration algorithm.

Step 0. Let $E^{(0)} = E_c$, $\mathbf{x}^{(0)} = \mathbf{x}$, $\mathbf{w}^{(0)} = \mathbf{w}$ and let $F^{(0)}$ be an empty set. Set $i = 0$.

Step 1. Compute the isotonic regression of $\mathbf{x}^{(i)}$ with respect to $\bar{E}^{(i)}$, the first $(\#E^{(i)} - 1)$ pairs in $E^{(i)}$. We will now list two possible cases, (a) and (b).

(a). If $E_{\mathbf{w}^{(i)}}(\mathbf{x}^{(i)} | K_{E^{(i)}})$ is isotonic with respect to $E^{(i)}$, then we drop the last pair in $E^{(i)}$ as a redundant pair. Let $\mathbf{x}^{(i+1)} = \mathbf{x}^{(i)}$, $E^{(i+1)} = \bar{E}^{(i)}$ and $F^{(i+1)} = F^{(i)}$. If $E^{(i+1)}$ contains only trees, we obtain a consistent partial ordering corresponding to $E_t = E^{(i+1)}$. Otherwise, set $i = i + 1$. Go to Step 1.

(b). If $E_{\mathbf{w}^{(i)}}(\mathbf{x}^{(i)} | K_{E^{(i)}})$ is not isotonic with respect to $E^{(i)}$, then we replace $x_{i_q}^{(i)}$ and $x_{j_q}^{(i)}$, as well as their weights $w_{i_q}^{(i)}$ and $w_{j_q}^{(i)}$ by $(w_{i_q}^{(i)} x_{i_q}^{(i)} + w_{j_q}^{(i)} x_{j_q}^{(i)}) / (w_{i_q}^{(i)} + w_{j_q}^{(i)})$ and $(w_{i_q}^{(i)} + w_{j_q}^{(i)})$ respectively, where $(i_q^{(i)}, j_q^{(i)})$ is the last element in $E^{(i)}$ while the remaining components of $\mathbf{x}^{(i+1)}$ and $\mathbf{w}^{(i+1)}$ stay the same as $\mathbf{x}^{(i)}$ and $\mathbf{w}^{(i)}$. Let $F^{(i+1)} = F^{(i)} \cup \{(i_q^{(i)}, j_q^{(i)})\}$ and let $E^{(i+1)} = \bar{E}^{(i)}$ but replacing $i_q^{(i)}$ and $j_q^{(i)}$ by a single index in $\bar{E}^{(i)}$. Delete all redundant pairs from $E^{(i+1)}$ (i.e. (i, j) is redundant if $(i, s) \in E^{(i)}$ and $(s, j) \in E^{(i)}$). If $E^{(i+1)}$ contains only trees, we obtain a consistent partial ordering corresponding to $E_t = E^{(i+1)} \cup F^{(i+1)}$. Otherwise, set $i = i + 1$. Go to Step 1.

Since $\#E^{(i)}$ ($i \geq 1$) decreases, when the number of each disconnected subset of

$E^{(i)}$ is not greater than 3, $E^{(i)}$ is sure to contain only trees. So there always exists a consistent partial ordering, which contains only trees, corresponding to $E_t = E^{(i)}$ such that $E_{\mathbf{w}}(\mathbf{x}|K_{E_c}) = E_{\mathbf{w}}(\mathbf{x}|K_{E_t})$. \square

The following example illustrates the iteration algorithm in Theorem 3.5. Using it, we will finally obtain a consistent partial ordering $E_t \subset E_c$, which contains only trees, such that $E_{\mathbf{w}}(\mathbf{x}|K_{E_c}) = E_{\mathbf{w}}(\mathbf{x}|K_{E_t})$.

Example 3.2. We will consider the simple loop. This partial ordering is the simplest ordering among those with cycles. We first order E as $\{(1, 2), (1, 3), (3, 4), (2, 4)\}$.

If the data $\mathbf{x} = (6, -2, -1, 2)$ with the equal weights $\mathbf{w} = (1, 1, 1, 1)$, then the isotonic regression with respect to the first three elements in E is $(1, 1, 1, 2)$. For the last element $(2, 4)$ in E , we found that $E_{\mathbf{w}^{(0)}}(\mathbf{x}^{(0)}|K_{\bar{E}^{(0)}})_2 = 1 \leq 2 = E_{\mathbf{w}^{(0)}}(\mathbf{x}^{(0)}|K_{\bar{E}^{(0)}})_4$. This is Case (a). So $(2, 4)$ is redundant. We now delete it and obtain the partial ordering $E_t = \{(1, 2), (2, 3), (3, 4)\}$ containing only a tree. The isotonic regression $\mathbf{x}^* = (1, 1, 1, 2)$ of $\mathbf{x} = (6, -2, -1, 2)$ with the equal weights corresponding to E is that one corresponding to $E_t = \{(1, 2), (1, 3), (3, 4)\}$.

If the data $\mathbf{x} = (2, 4, -1, -5)$ with the equal weights $\mathbf{w} = (1, 1, 1, 1)$, then the isotonic regression with respect to the first three elements in E is $(-4/3, 4, -4/3, -4/3)$. Now we consider the last element $(2, 4)$ in E . Since $E_{\mathbf{w}^{(0)}}(\mathbf{x}^{(0)}|K_{\bar{E}^{(0)}})_2 = 4 \not\leq -4/3 = E_{\mathbf{w}^{(0)}}(\mathbf{x}^{(0)}|K_{\bar{E}^{(0)}})_4$, this is Case (b). We obtain the reduced problem by pooling $x_2^{(0)}$ and $x_4^{(0)}$ ordered by $(2, 4)$ so that $\mathbf{x}^{(1)} = (2, -1, -1/2)$, $\mathbf{w}^{(1)} = (1, 1, 2)$,

$E^{(1)} = \{(1, 4^*), (1, 3), (3, 4^*)\}$ and $F^{(1)} = \{(2, 4)\}$. However, $(1, 4^*)$ is a redundant pair in $E^{(1)}$ and hence $E^{(1)} = \{(1, 3), (3, 4^*)\}$. The remainder partial ordering corresponding to $E_t = \{(1, 3), (3, 4), (2, 4)\}$ contains only a tree. The isotonic regression $\mathbf{x}^* = (0, 0, 0, 0)$ of $\mathbf{x} = (2, 4, -1, -5)$ with the equal weights corresponding to E is that one corresponding to $E_t = \{(1, 3), (3, 4), (2, 4)\}$.

4. THE GENERALIZED ISOTONIC REGRESSION PROBLEM

Barlow and Brunk (1972) formulated a generalization of the isotonic regression problem. The solution is closely associated with the isotonic regression. The authors provided two different approaches in proving their Theorem 3.1 for the generalized isotonic regression problem using the concavity of function and the Fenchel duality theorem. In this section, we will use the Kuhn-Tucker equivalence theorem to prove their theorem and provide a new approach for solving some related problems in Section 5. Also, we will show some important results for isotonic regression using the Kuhn-Tucker equivalence theorem. As will be seen, some proofs are much simpler using the Kuhn-Tucker equivalence theorem than those approaches in the literature.

Let $\mathbf{g} = (g_1, \dots, g_k)$ and $\mathbf{w} = (w_1, \dots, w_k) > 0$ be given functions on $\Omega = \{1, 2, \dots, k\}$. Let Φ be a proper convex function on \mathbb{R} .

The generalized isotonic regression problem due to Barlow and Brunk (1972) is to

$$\begin{aligned} & \text{Minimize } \sum_{i=1}^k [\Phi(x_i) - g_i x_i] w_i, \\ & \text{subject to } x_i \leq x_j \text{ when } i \prec j, \end{aligned} \tag{4.1}$$

where the binary relation \prec is a specified partial ordering on Ω .

If $\Phi(x) = x^2/2$, then the generalized problem (4.1) is formally equivalent to the isotonic regression problem, since

$$\sum_{i=1}^k \left[\frac{x_i^2}{2} - g_i x_i \right] w_i = \frac{1}{2} \sum_{i=1}^k [x_i - g_i]^2 w_i - \frac{1}{2} \sum_{i=1}^k g_i^2 w_i.$$

The following theorem due to Barlow and Brunk (1972) provides the solutions to many problems of restricted maximum likelihood estimation for ordered parameters of distributions such as multinomial, Poisson, normal, beta, and gamma distributions. We will verify this theorem using the Kuhn-Tucker equivalence theorem.

Theorem 4.1. (Barlow and Brunk). Let Φ be a proper convex function on R , φ a finite determination of its derivative and let $\varphi^{-1}(t) = \inf\{x : \varphi(x) > t\}$. Suppose K is a convex cone of isotonic functions on Ω having range in R . Suppose also that the range of \mathbf{g} is in the effective domain of φ^{-1} where \mathbf{g} is a given function on Ω . Then the solution \mathbf{x}^* to the generalized problem (4.1) is given by $\varphi^{-1}(E_{\mathbf{w}}(\mathbf{g}|K))$, where $E_{\mathbf{w}}(\mathbf{g}|K)$ is the isotonic regression of \mathbf{g} with respect to the weights \mathbf{w} . The solution is unique if Φ is strictly convex.

Proof. The generalized problem (4.1) is equivalent to

$$\begin{aligned} \text{Maximize} \quad & - \sum_{i=1}^k [\Phi(x_i) - g_i x_i] w_i, \\ \text{subject to} \quad & x_i \leq x_j \text{ when } i < j. \end{aligned} \tag{4.2}$$

since Φ is a proper convex function. Using the Kuhn-Tucker equivalence theorem, we construct the function

$$\Psi(\mathbf{x}, \lambda) = - \sum_{i=1}^k [\Phi(x_i) - g_i x_i] w_i + \sum_{(i,j) \in E} \lambda_{ij} (x_j - x_i).$$

Then the solution to problem (4.2) is \mathbf{x}^* if and only if there exists a vector λ such

that

$$(4.3) \left\{ \begin{array}{l} \frac{\partial \Psi}{\partial x_i} \Big|_{\mathbf{x}^*} = [g_i - \varphi(x_i^*)]w_i - \sum_{j \in S_i} \lambda_{ij} + \sum_{j \in P_i} \lambda_{ji} = 0, \quad i = 1, \dots, k \quad (4.3)^* \\ \frac{\partial \Psi}{\partial \lambda_{ij}} \Big|_{\mathbf{x}^*} = x_j^* - x_i^* \geq 0, \quad (i, j) \in E \\ \lambda_{ij} \left(\frac{\partial \Psi}{\partial \lambda_{ij}} \Big|_{\mathbf{x}^*} \right) = \lambda_{ij}(x_j^* - x_i^*) = 0, \quad (i, j) \in E \\ \lambda_{ij} \geq 0, \quad (i, j) \in E \end{array} \right.$$

where

$$S_i = \{j : (i, j) \in E\},$$

$$P_i = \{j : (j, i) \in E\}, \quad i \in \Omega.$$

From the Kuhn-Tucker conditions (4.3), we have $x_j^* - x_i^* \geq 0$ if $(i, j) \in E$.

And

$$\begin{aligned} & \sum_{i=1}^k [g_i - \varphi(x_i^*)][\varphi(x_i^*)]w_i \\ &= \sum_{i=1}^k \left(\sum_{j \in S_i} \lambda_{ij} - \sum_{j \in P_i} \lambda_{ji} \right) [\varphi(x_i^*)] \\ &= - \sum_{(i,j) \in E} \lambda_{ij} (\varphi(x_j^*) - \varphi(x_i^*)) \\ &= 0, \end{aligned}$$

since $\varphi(x)$ is nondecreasing in x and $\lambda_{ij}(x_j^* - x_i^*) = 0$ if $(i, j) \in E$.

For any $\mathbf{x}' \in K$,

$$\begin{aligned}
& \sum_{i=1}^k [g_i - \varphi(x_i^*)] x'_i w_i \\
&= \sum_{i=1}^k \left(\sum_{j \in S_i} \lambda_{ij} - \sum_{j \in P_i} \lambda_{ji} \right) x'_i \\
&= - \sum_{(i,j) \in E} \lambda_{ij} (x'_j - x'_i) \\
&\leq 0,
\end{aligned}$$

since $\lambda_{ij} \geq 0$ and $x'_j \geq x'_i$ if $(i, j) \in E$. By Theorem 2.1, we have

$$\mathbf{x}^* = \varphi^{-1}(E_{\mathbf{w}}(\mathbf{g}|K)).$$

If Φ is strictly convex, the solution is unique since K is a closed convex cone. \square

Note that the parameters λ_{ij} 's, $(i, j) \in E$, are involved in the Kuhn-Tucker conditions (4.3). Next we will consider the values of these λ_{ij} 's. From Theorem 4.1, we know that $\mathbf{x}^* = \varphi^{-1}(E_{\mathbf{w}}(\mathbf{g}|K))$ is the unique solution to problem (4.1) if Φ is strictly convex. In this case, we will present the unique values of the parameters λ_{ij} 's, $(i, j) \in E$ if the partial ordering corresponding to E contains only trees. However, the values of these λ_{ij} 's may not be unique for the partial ordering with cycles.

Case One. The partial ordering corresponding to E contains only $m(\geq 1)$ disconnected trees. It suffices to consider that the partial ordering is a tree, i.e. $m = 1$. Then $\#E = k - 1$. And apparently, $\pm \mathbf{1}_k = \pm(1, \dots, 1)'$ belong to K .

Suppose

$$\mathbf{d}' = ([g_1 - \varphi(x_1^*)]w_1, \dots, [g_k - \varphi(x_k^*)]w_k),$$

$$\mathbf{e}' = (\varphi(x_1^*), \dots, \varphi(x_k^*)).$$

Since $\mathbf{x}^* = \varphi^{-1}[E_{\mathbf{w}}(\mathbf{g}|K)]$, by Theorem 2.1, $\varphi(\mathbf{x}^*) = E_{\mathbf{w}}(\mathbf{g}|K)$ satisfies

$$\left\{ \begin{array}{l} \varphi(x_j^*) \geq \varphi(x_i^*), \quad (i, j) \in E \end{array} \right. \quad (4.4)$$

$$\left\{ \begin{array}{l} \sum_{i=1}^k [g_i - \varphi(x_i^*)][\varphi(x_i^*)]w_i = \langle \mathbf{d}, \mathbf{e} \rangle = 0 \end{array} \right. \quad (4.5)$$

$$\left\{ \begin{array}{l} \sum_{i=1}^k [g_i - \varphi(x_i^*)]x'_i w_i = \langle \mathbf{d}, \mathbf{x}' \rangle \leq 0, \quad \forall \mathbf{x}' \in K. \end{array} \right. \quad (4.6)$$

By Condition (4.6), we can obtain

$$\langle \mathbf{d}, \mathbf{1}_k \rangle = 0.$$

Note that we have obtained the expressions of \mathbf{b}_{ij} 's, $(i, j) \in E$, in Theorem 3.4.

Clearly, all \mathbf{b}_{ij} 's are in K . And by Condition (4.6),

$$\langle \mathbf{d}, \mathbf{b}_{ij} \rangle \leq 0, \quad (i, j) \in E.$$

Observe that the equations (4.3)* are equivalent to

$$\left\{ \begin{array}{l} \langle \mathbf{d}, \mathbf{1}_k \rangle = 0 \\ \langle \mathbf{d}, \mathbf{b}_{ij} \rangle + \lambda_{ij} = 0, \quad (i, j) \in E. \end{array} \right.$$

Let $\lambda_{ij} = -\langle \mathbf{d}, \mathbf{b}_{ij} \rangle \geq 0$, $(i, j) \in E$. Thus, \mathbf{x}^* and λ satisfy the equations (4.3)* and $x_j^* \geq x_i^*$, $\lambda_{ij} \geq 0$, $(i, j) \in E$. Since $\varphi(x)$ is increasing in x , by Theorem 3.3 and 3.4, the vector \mathbf{e} can be expressed as

$$\mathbf{e} = \gamma \mathbf{1}_k + \sum_{(i,j) \in E} (\lambda_{ij} - \lambda_{ji}) \mathbf{b}_{ij}.$$

By Condition (4.5), we have

$$\begin{aligned}
0 &= \sum_{i=1}^k [g_i - \varphi(x_i^*)][\varphi(x_i^*)]w_i = \langle \mathbf{d}, \mathbf{e} \rangle \\
&= \langle \mathbf{d}, \gamma \mathbf{1}_k + \sum_{(i,j) \in E} (e_j - e_i) \mathbf{b}_{ij} \rangle \\
&= \gamma \langle \mathbf{d}, \mathbf{1}_k \rangle + \sum_{(i,j) \in E} (e_j - e_i) \langle \mathbf{d}, \mathbf{b}_{ij} \rangle \\
&= - \sum_{(i,j) \in E} \lambda_{ij} (e_j - e_i).
\end{aligned}$$

So

$$\lambda_{ij}(e_j - e_i) = 0 \text{ for each } (i, j) \in E$$

since $\lambda_{ij} \geq 0$ and $e_j - e_i \geq 0$.

Further,

$$\lambda_{ij}(x_j^* - x_i^*) = 0, \quad (i, j) \in E$$

since $\varphi(x)$ is increasing in x . Thus, \mathbf{x}^* and λ_{ij} 's satisfy the Kuhn-Tucker conditions (4.3).

Case Two. E contains cycles. By Theorem 3.5, there exists a consistent partial ordering corresponding to E_t , which contains only trees, such that

$$E_{\mathbf{w}}(\mathbf{g}|K_E) = E_{\mathbf{w}}(\mathbf{g}|K_{E_t}) = \varphi(\mathbf{x}^*).$$

By *Case one*, \mathbf{x}^* and λ_{ij} 's, $(i, j) \in E_t$, satisfy the Kuhn-Tucker conditions (4.3) corresponding to E_t . However, we compared the Kuhn-Tucker conditions (4.3) corresponding to E with E_t and found that the difference between them is that there exist some λ_{ij} 's corresponding to the redundant pairs in E in the course of the iteration

algorithm of Theorem 3.5. Therefore, we set all such λ_{ij} 's equal to 0. Then \mathbf{x}^* and λ_{ij} 's, $(i, j) \in E$, satisfy the Kuhn-Tucker conditions (4.3) corresponding to E .

The calculation of $E_{\mathbf{w}}(\mathbf{g}|K_E)$, given \mathbf{g} , the weights \mathbf{w} , and the partial ordering on Ω corresponding to E , can be accomplished via quadratic programming. An extensive literature on methods of computing quadratic programming solutions for such problems exists. The problem of computing the isotonic regression is special and a number of algorithms have been proposed for this specific problem. The graphical representation and the pool-adjacent-violators algorithm (PAVA) discussed in Section 6 are very elegant algorithms, but they apply only to simple order restrictions. The algorithms presented here all involve averaging \mathbf{g} over suitably selected subsets of Ω ; the term *amalgamation of means* has frequently been used in connection with such computations.

The *Minimum Lower Sets* algorithm is given in Brunk, Ewing and Utz (1957) and in Brunk (1955). The iterative algorithm for the matrix partial ordering is developed in Dykstra and Robertson (1982). This type of iterative algorithm has been extended to a large number of restricted optimization problems (cf. Dykstra (1983) and Lee (1983)). We will prove the *Minimum Lower Sets* algorithm by using the Kuhn-Tucker equivalence theorem. Our proof is much simpler than Theorem 2.7 due to Barlow's et al. (1972).

To simplify the notation, throughout the remainder of the section L and U will only be used to denote lower sets and upper sets respectively.

Theorem 4.2. (Minimum Lower Sets). Suppose that $a_1^* < a_2^* < \dots < a_l^*$ are the l distinct values taken by the solution \mathbf{x}^* to problem (4.1). Let $B_m = \{j : x_j^* = a_m^*\}$ and $L_m = \bigcup_{i=1}^m B_i$, $m = 1, \dots, l$. Then for $m = 1, \dots, l$,

$$Av(B_m) = \min_L Av(L - L_{m-1}),$$

$$Av(B_m) < Av(L - L_{m-1}), \quad \forall B_m \subset L - L_{m-1},$$

where for $\forall B \subseteq \Omega$,

$$Av(B) = \frac{\sum_{i \in B} g_i w_i}{\sum_{i \in B} w_i}.$$

Proof. By the construction of B_m , $m = 1, \dots, l$, the Kuhn-Tucker conditions (4.3) imply

$$\lambda_{ij} \begin{cases} = 0, & (i, j) \in E \text{ and } i \in B_m, j \notin B_m \text{ or } j \in B_m, i \notin B_m, \quad m = 1, \dots, l, \\ \geq 0, & o/w. \end{cases}$$

And

$$\varphi(x_i^*) = Av(B_m), \quad i \in B_m, \quad m = 1, \dots, l.$$

It is straightforward to show that L_m , $m = 1, \dots, l$, are lower sets. For any $L - L_{m-1}$, let $G_{L-L_{m-1}} = \{(i, j) \in E : i \in L - L_{m-1}, j \notin L - L_{m-1}\}$.

If $L - L_{m-1} \neq \emptyset$, since $\lambda_{ij} = 0$, $(i, j) \in E$, $i \in B_m$, $j \notin B_m$ or $j \in B_m$, $i \notin$

B_m , $m = 1, \dots, l$ and $\lambda_{ij} \geq 0$, o/w , by the equations (4.3)*, We have

$$\begin{aligned}
& \sum_{i \in L-L_{m-1}} [g_i - \varphi(x_i^*)] w_i - \sum_{(i,j) \in \tilde{G}_{L-L_{m-1}}} \lambda_{ij} = 0 \quad (4.7) \\
\Rightarrow & \sum_{i \in L-L_{m-1}} [g_i - \varphi(x_i^*)] w_i \geq 0 \\
\Rightarrow & \sum_{i \in L-L_{m-1}} g_i w_i \geq \sum_{i \in L-L_{m-1}} \varphi(x_i^*) w_i \geq Av(B_m) \sum_{i \in L-L_{m-1}} w_i \\
\Rightarrow & Av(L - L_{m-1}) \geq Av(B_m).
\end{aligned}$$

If $L - L_{m-1} \supset B_m$, from (4.7) we have

$$\begin{aligned}
& \sum_{i \in L-L_{m-1}} g_i w_i \\
& \geq \sum_{i \in L-L_{m-1}} \varphi(x_i^*) w_i \\
& = \sum_{i \in B_m} \varphi(x_i^*) w_i + \sum_{i \in L-L_{m-1}-B_m} \varphi(x_i^*) w_i \\
& > Av(B_m) \sum_{i \in L-L_{m-1}} w_i.
\end{aligned}$$

Therefore, $Av(L - L_{m-1}) > Av(B_m)$. □

The *Max-Min formulas* for isotonic regression are the important results in the theory of isotonic regression. We can also use the Kuhn-Tucker equivalence theorem to prove these results.

Theorem 4.3. (Max-Min formulas). The solution \mathbf{x}^* to problem (4.1) is given

by

$$\begin{aligned}
x_i^* &= \max_{U:i \in U} \min_{L:i \in L} \varphi^{-1}(Av(L \cup U)) \\
&= \max_{U:i \in U} \min_{L:L \cup U \neq \emptyset} \varphi^{-1}(Av(L \cup U)) \\
&= \min_{L:i \in L} \max_{U:i \in U} \varphi^{-1}(Av(L \cup U)) \\
&= \min_{L:i \in L} \max_{U:L \cup U \neq \emptyset} \varphi^{-1}(Av(L \cup U)).
\end{aligned}$$

Proof. We use the notations in Theorem 4.2. For each $i \in \Omega$, there exists $B_m = L_m - L_{m-1}$ such that $i \in B_m$ and

$$\varphi(x_i^*) = Av(B_m)$$

by the proof of Theorem 4.2. Now we consider

$$\max_{U:i \in U} \min_{L:i \in L} Av(L \cap U).$$

By Theorem 4.2, taking $U_{m-1} = L_{m-1}^c$, then we have $Av(B_m) = \min_{L:i \in L} Av(L \cap U_{m-1})$.

It suffices to show that

$$Av(B_m) = \max_{U:i \in U} \min_{L:i \in L} Av(L \cap U).$$

For any $U : i \in U$, we consider $L_m \cap U$. There exist some B_{i_1}, \dots, B_{i_r} , $i_1, \dots, i_r \in \{1, \dots, m\}$, such that $U \cap B_{i_g} \neq \emptyset$, $g = 1, \dots, r$, and $L_m \cap U = \bigcup_{g=1}^r (U \cap B_{i_g})$.

For any $U \cap B_{i_g} \neq \emptyset$, $g = 1, \dots, r$, set

$$H_{U \cap B_{i_g}} = \{(i, j) \in E : i \in B_{i_g} - B_{i_g} \cap U, j \in B_{i_g} \cap U\}.$$

Since $\lambda_{ij} = 0$, $(i, j) \in E$, $i \in B_m$, $j \notin B_m$ or $j \in B_m$, $i \notin B_m$, $m = 1, \dots, l$,

$$\begin{aligned}
& \sum_{i \in U \cap B_{i_g}} \{[g_i - \varphi(x_i^*)]w_i - \sum_{j \in S_i} \lambda_{ij} + \sum_{j \in P_i} \lambda_{ji}\} = 0 \\
\Rightarrow & \sum_{i \in U \cap B_{i_g}} [g_i - \varphi(x_i^*)]w_i + \sum_{i \in U \cap B_{i_g}} (-\sum_{j \in S_i} \lambda_{ij} + \sum_{j \in P_i} \lambda_{ji}) = 0 \\
\Rightarrow & \sum_{i \in U \cap B_{i_g}} [g_i - \varphi(x_i^*)]w_i + \sum_{(i,j) \in H_{U \cap B_{i_g}}} \lambda_{ij} = 0 \\
\Rightarrow & \sum_{i \in U \cap B_{i_g}} [g_i - \varphi(x_i^*)]w_i \leq 0 \\
\Rightarrow & Av(U \cap B_{i_g}) \leq Av(B_{i_g}) \leq Av(B_m).
\end{aligned}$$

By Theorem 4.2, $Av(B_i) < Av(B_j)$, $i < j$. Therefore,

$$\begin{aligned}
& Av(L_m \cap U) = Av\left(\bigcup_{g=1}^r (U \cap B_{i_g})\right) \leq Av(B_m) \\
\Rightarrow & \min_{L: i \in L} Av(L \cap U) \leq Av(B_m) \\
\Rightarrow & \max_{U: i \in U} \min_{L: i \in L} Av(L \cap U) = Av(B_m) = \varphi(x_i^*) \\
\Rightarrow & x_i^* = \max_{U: i \in U} \min_{L: i \in L} \varphi^{-1}(Av(L \cap U)).
\end{aligned}$$

The other parts of the theorem are argued similarly. □

5. SOME RELATED PROBLEMS

In the beginning, I tried to use the Kuhn-Tucker equivalence theorem to solve restricted Maximum Likelihood Estimation (MLE) for simply ordered binomial parameters. And then I generalized this approach to restricted MLE for partially ordered parameters of exponential families of distributions. This provides a very different way to prove Theorem 4.1. In this section, we will discuss four problems using the Kuhn-Tucker equivalence theorem.

5.1. Exponential Families

Robertson et al. (1988) studied the order restricted MLE for exponential families of distributions. They considered a regular one-parameter exponential family of distributions defined by probability densities of the form

$$f(y; \theta, \tau) = \exp\{p_1(\theta)p_2(\tau)K(y; \tau) + S(y; \tau) + q(\theta, \tau)\}$$

for $y \in A$, $\theta \in (\theta_1, \theta_2)$, $\tau \in T$.

Here the parameter τ is thought of as a nuisance parameter and it was treated as a known one. Their Theorem 1.5.2 requires three regularity conditions.

For an application of the Kuhn-Tucker equivalence theorem, we consider the familiar one-parameter exponential family of distributions

$$f(x; \theta) = \exp\{c(\theta)T(x) + d(\theta) + S(x)\}I_A(x), \quad (5.1)$$

where $\theta \in (\theta_1, \theta_2)$.

We require the following two regularity conditions for θ :

1. $c(\theta)$ and $d(\theta)$ both have continuous second derivatives on (θ_1, θ_2) .
2. $c'(\theta) > 0$ for all $\theta \in (\theta_1, \theta_2)$.

Under these regularity conditions, the density function (5.1) can be expressed in the natural form

$$\begin{aligned} f(x; \theta) &= f(x; \eta) \\ &= \exp\{\eta T(x) + d_0(\eta) + S(x)\} I_A(x), \end{aligned} \tag{5.2}$$

where $\eta = c(\theta)$ is increasing in θ , $d_0(\eta) = d(c^{-1}(\eta))$. And by Theorem 2.3.2 (cf. Bickel and Doksum (1977)), we have

$$\begin{aligned} E(T(X)) &= -d'_0(\eta), \\ \text{Var}(T(X)) &= -d''_0(\eta) > 0. \end{aligned}$$

Suppose one has independent random samples from k populations belonging to the above one-parameter exponential family with the sample sizes denoted by n_i and the observations by X_{ij} , i.e. X_{ij} has density $f(x; \theta_i)$ for $j = 1, \dots, n_i$ and $i = 1, \dots, k$. If one has prior information that θ is isotonic with respect to a specified partial ordering, an estimate of θ possessing the property is desired. The log likelihood function is

$$\begin{aligned} l(\theta) &= \sum_{i=1}^k \{n_i [c(\theta_i) \bar{T}_i + d(\theta_i)] + \sum_{j=1}^{n_i} S(x_{ij})\} \\ &= \sum_{i=1}^k n_i [\eta_i \bar{T}_i + d_0(\eta_i)] + \sum_{i=1}^k \sum_{j=1}^{n_i} S(x_{ij}), \end{aligned} \tag{5.3}$$

where $\bar{T}_i = \sum_{j=1}^{n_i} T(x_{ij})/n_i$.

Clearly, the restricted MLE θ^* of θ which maximizes

$$\sum_{i=1}^k n_i [c(\theta_i) \bar{T}_i + d(\theta_i)]$$

is equivalent to the restricted MLE η^* of η which maximizes

$$\sum_{i=1}^k n_i [\eta_i \bar{T}_i + d_0(\eta_i)]$$

since $\eta = c(\theta)$ is increasing in θ .

Notice that $n_i [\eta_i \bar{T}_i + d_0(\eta_i)]$ is a strictly concave function of η_i since

$$\begin{aligned} \{n_i [\eta_i \bar{T}_i + d_0(\eta_i)]\}'' &= n_i d_0''(\eta_i) \\ &= -n_i \text{Var}(T_i(X)) < 0. \end{aligned}$$

Using the Kuhn-Tucker equivalence theorem and a similar proof of Theorem 4.1, we can first obtain the unique order restricted MLE η^* of $\eta : -d_0''^{-1}(E_{\mathbf{n}}(\bar{\mathbf{T}}|K))$, where

$$\mathbf{n} = (n_1, \dots, n_k), \quad \bar{\mathbf{T}} = (\bar{T}_1, \dots, \bar{T}_k).$$

Then $\theta^* = c^{-1}(\eta^*)$.

Remark 5.1.

1. Compared to the regularity conditions of Theorem 1.5.2 (cf. Robertson et al. (1988)), our conditions are essentially its first two conditions.

2. In Robertson et al. (1988), the third condition of Theorem 1.5.2 was used for a transformation. Thus, they use Theorem 1.5.1 to obtain the restricted MLE of θ . On the other hand, our transformation is more natural and simpler. We directly use the properties of the exponential family of distributions and the concavity of $d_0(\eta)$.

5.2. A Dykstra-Lee Theorem

In order to discuss the multinomial estimation procedures for isotonic cones, Dykstra and Lee (1991) showed a generalization of Theorem 4.1 in terms of a saddlepoint of Lagrangian function.

Theorem 5.1. (Dykstra and Lee). Under the conditions of Theorem 4.1, the solution to the problem

$$\begin{aligned} & \text{Minimize } \sum_{i=1}^k [\Phi(x_i) - g_i x_i] w_i, \\ & \text{subject to } x_i \leq x_j \text{ when } i \prec j, \\ & \text{and } \sum_{i=1}^k a_i x_i = c \end{aligned} \tag{5.4}$$

is given by $\varphi^{-1}[E_{\mathbf{w}}(\mathbf{g} + \lambda_0 \mathbf{a}/\mathbf{w}|K)]$ (where vector multiplication and division are done coordinatewise) provided that

$$\sum_{i=1}^k a_i \varphi^{-1}[E_{\mathbf{w}}(\mathbf{g} + \lambda_0 \mathbf{a}/\mathbf{w}|K)_i] = c.$$

Proof. Let

$$\begin{aligned} \Psi(\mathbf{x}, \lambda) &= - \sum_{i=1}^k [\Phi(x_i) - g_i x_i] w_i + \sum_{(i,j) \in E} \lambda_{ij} (x_j - x_i) \\ &+ \lambda_0 \left(\sum_{i=1}^k a_i x_i - c \right). \end{aligned} \tag{5.5}$$

By the Kuhn-Tucker equivalence theorem, the solution to problem (5.4) is \mathbf{x}^* if and only if there exists a vector λ such that

$$(5.6) \left\{ \begin{array}{l} [(g_i + \lambda_0 a_i / w_i) - \phi(x_i^*)] w_i - \sum_{j \in S_i} \lambda_{ij} + \sum_{j \in P_i} \lambda_{ji} = 0, \quad i = 1, \dots, k \\ x_j^* - x_i^* \geq 0, \quad (i, j) \in E \\ \lambda_{ij} (x_j^* - x_i^*) = 0, \quad (i, j) \in E \\ \lambda_{ij} \geq 0, \quad (i, j) \in E \\ \sum_{i=1}^k a_i x_i^* = c. \end{array} \right. \tag{5.6}^*$$

Similarly to the proof of Theorem 4.1, from (5.6)*, we conclude that $\mathbf{x}^* = \varphi^{-1}\{E_{\mathbf{w}}(\mathbf{g} + \lambda_0 \mathbf{a}/\mathbf{w}|K)\}$. Then the result of Theorem 5.1 follows. \square

5.3. I-projection Problems

Dykstra (1985) dealt with the I-projection problems using the Fenchel duality theorem. Consider the problem

$$\text{Minimize } \sum_{i=1}^k p_i \ln\left(\frac{p_i}{r_i}\right), \quad (5.7)$$

subject to $p_i \leq p_j$ when $i \prec j$,

where \mathbf{p} and \mathbf{r} are two probability vectors and \mathbf{r} is given. The binary relation \prec is a specified partial ordering on $\Omega = \{1, 2, \dots, k\}$.

Let P be the collection of probability vectors.

The solution to problem (5.7) is often referred to as an I-projection of \mathbf{r} onto $K \cap P$ and is important in a wide variety of problems. In particular, minimization problems of the form (5.7) play a key role in the information-theoretic approach to statistics (e.g. Kullback (1959), and Good (1963)) and also occur in other areas such as the theory of large deviations (Sanov (1957)) and maximization of entropy (Rao (1965), and Jaynes (1957)). This type of problem also occurs in maximum likelihood estimation problems for log-linear models.

Note that $-[p \ln(p/r)]$ is a concave function of p over the hyperplane $\sum_{i=1}^k p_i = 1$. Directly using the Kuhn-Tucker equivalence theorem and a similar proof of Theorem 5.1, the solution to problem (5.7) is given by $\mathbf{p}^* = \exp \{E[\ln \mathbf{r} + (\lambda_0 - 1)\mathbf{1}_k|K]\}$

provided that

$$\sum_{i=1}^k \exp \{E[\ln \mathbf{r} + (\lambda_0 - 1)\mathbf{1}_k|K]\} = 1, \quad (5.8)$$

where $\mathbf{1}'_k = (1, 1, \dots, 1)$. And $E(\mathbf{x}|K)$ is the isotonic regression of \mathbf{x} with respect to the equal weights.

However, we can show

$$E[\ln \mathbf{r} + (\lambda_0 - 1)\mathbf{1}_k|K] = E[\ln \mathbf{r}|K] + (\lambda_0 - 1)\mathbf{1}_k.$$

In fact, it suffices to check the necessary and sufficient conditions of Theorem 2.1.

First, since $E[\ln \mathbf{r}|K] \in K$, for $\forall (i, j) \in E$,

$$E[\ln \mathbf{r}|K]_i \leq E[\ln \mathbf{r}|K]_j$$

implies that

$$E[\ln \mathbf{r}|K]_i + \lambda_0 - 1 \leq E[\ln \mathbf{r}|K]_j + \lambda_0 - 1.$$

Thus, $E[\ln \mathbf{r}|K] + (\lambda_0 - 1)\mathbf{1}_k \in K$.

Secondly,

$$\begin{aligned} & \sum_{i=1}^k [(\ln r_i + \lambda_0 - 1) - (E[\ln \mathbf{r}|K]_i + \lambda_0 - 1)][E[\ln \mathbf{r}|K]_i + \lambda_0 - 1] \\ &= \sum_{i=1}^k [\ln r_i - E[\ln \mathbf{r}|K]_i][E[\ln \mathbf{r}|K]_i + \lambda_0 - 1] \\ &= \sum_{i=1}^k [\ln r_i - E[\ln \mathbf{r}|K]_i][E[\ln \mathbf{r}|K]_i + (\lambda_0 - 1) \sum_{i=1}^k [\ln r_i - E[\ln \mathbf{r}|K]_i]] \\ &= 0 \end{aligned}$$

by the properties of the isotonic regression:

$$\sum_{i=1}^k [\ln r_i - E[\ln \mathbf{r}|K]_i] E[\ln \mathbf{r}|K]_i = 0,$$

$$\sum_{i=1}^k \ln r_i = \sum_{i=1}^k E[\ln \mathbf{r}|K]_i.$$

Finally, for $\forall \mathbf{p}' \in K$,

$$\begin{aligned} & \sum_{i=1}^k [(\ln r_i + \lambda_0 - 1) - (E[\ln \mathbf{r}|K]_i + \lambda_0 - 1)]p'_i \\ &= \sum_{i=1}^k [\ln r_i - E[\ln \mathbf{r}|K]_i]p'_i \\ &\leq 0. \end{aligned}$$

Solving the equation (5.8), we obtain the solution \mathbf{p}^* to the problem (5.7) given by

$$\exp[E(\ln \mathbf{r}|K)] / \sum_{i=1}^k \exp[E(\ln \mathbf{r}|K)_i]. \quad (5.9)$$

Remark 5.2. Cressie and Read (1984) defined the divergence of the PV \mathbf{r} with respect to the PV \mathbf{p} of order λ as

$$I^\lambda(\mathbf{r} : \mathbf{p}) = \frac{1}{\lambda(\lambda + 1)} \sum_{i=1}^k r_i \left[\left(\frac{r_i}{p_i} \right)^\lambda - 1 \right]. \quad (5.10)$$

The I-projection problem (5.7) is a special case of (5.10) when $\lambda = -1$. Dykstra and Lee (1991) gave the general results for $\lambda \in R$. They used Theorem 5.1 and some properties of the isotonic regression to obtain those results. Although we only discussed the case of $\lambda = -1$ here, in fact, we may use the Kuhn-Tucker equivalence theorem to show their general results. The proof is similar to that of problem (5.7)

but we still need those properties of the isotonic regression used by Dykstra and Lee (1991).

5.4. Maximum Likelihood Estimations of the Probabilities in k Binomial Populations

For $i = 1, \dots, k$, let n_i independent trials be made of an event with probability p_i , and suppose that the probabilities p_i are known to satisfy the inequalities $p_1 \geq p_2 \geq \dots \geq p_k$. Let a_i and b_i denote the number of successes and failures in the i -th trial, respectively, and \bar{p}_i the ratio a_i/n_i ($i = 1, 2, \dots, k$). Ayer et al. (1955) showed their Theorem 2.1 on the basis of the monotonic property of $p^a(1-p)^b$ with respect to p . Their theorem provides an important means of calculating the maximum likelihood estimate \mathbf{p}^* of $\mathbf{p} = (p_1, \dots, p_k)$ —the pool-adjacent-violators algorithm (PAVA) discussed in the next section. Here we will use the Kuhn-Tucker equivalence theorem to prove their theorem 2.1.

Theorem 5.2. (Ayer et al.). If \mathbf{p}^* is the restricted maximum likelihood estimate of \mathbf{p} , and if $p_i^* > p_{i+1}^*$ for some i , $1 \leq i \leq k-1$, then $\bar{p}_i \geq p_i^* > p_{i+1}^* \geq \bar{p}_{i+1}$. Also, $\bar{p}_1 \leq p_1^*$, and $\bar{p}_k \geq p_k^*$.

Proof. The log likelihood function for k populations is

$$l(\mathbf{p}) = \sum_{i=1}^k \{a_i \ln p_i + b_i \ln(1 - p_i)\} + \sum_{i=1}^k \ln \binom{a_i + b_i}{a_i}.$$

Since the second term in $l(\mathbf{p})$ does not involve \mathbf{p} , the restricted MLE \mathbf{p}^* solves the

reduced problem

$$\text{Maximize } \sum_{i=1}^k \{a_i \ln p_i + b_i \ln(1 - p_i)\},$$

$$\text{subject to } p_1 \geq p_2 \geq \cdots \geq p_k.$$

Since the function $\{a \ln p + b \ln(1 - p)\}$ is a concave function of p , by the Kuhn-

Tucker equivalence theorem, let

$$\Psi(\mathbf{p}, \lambda) = \sum_{i=1}^k \{a_i \ln p_i + b_i \ln(1 - p_i)\} + \sum_{i=1}^{k-1} \lambda_i (p_i - p_{i+1}).$$

Then \mathbf{p}^* is the restricted maximum likelihood estimate of \mathbf{p} if and only if there

exists a vector λ such that

$$\begin{cases} \frac{a_i}{p_i^*} - \frac{b_i}{1-p_i^*} + \lambda_i - \lambda_{i-1} = 0, & i = 1, \dots, k \\ \lambda_i (p_i^* - p_{i+1}^*) = 0, & i = 1, \dots, k-1 \\ p_1^* \geq p_2^* \geq \cdots \geq p_k^* \\ \lambda_i \geq 0, & i = 1, \dots, k-1 \end{cases} \quad (5.11)$$

with convention $0/0 = 0$ and $\lambda_0 = \lambda_k = 0$.

Assume that $\bar{p}_i < p_i^*$. From the Kuhn-Tucker conditions (5.11), we have

$$p_i^* > p_{i+1}^* \implies \lambda_i = 0 \implies \frac{a_i}{p_i^*} - \frac{b_i}{1-p_i^*} - \lambda_{i-1} = 0$$

$$\implies a_i(1 - p_i^*) - b_i p_i^* - \lambda_{i-1} p_i^*(1 - p_i^*) = 0$$

$$\implies a_i = (a_i + b_i)p_i^* + \lambda_{i-1} p_i^*(1 - p_i^*)$$

$$\implies \bar{p}_i = p_i^* + \lambda_{i-1} p_i^*(1 - p_i^*) / (a_i + b_i)$$

$$\implies \bar{p}_i - p_i^* = \lambda_{i-1} p_i^*(1 - p_i^*) / (a_i + b_i)$$

$$\implies \lambda_{i-1} < 0. \quad \text{Contradiction.}$$

Therefore, $\bar{p}_i \geq p_i^*$. Similarly, we can prove the other parts of the theorem. \square

6. ON THE SIMPLE ORDERING

Here we discuss some results on a particular case of partial orderings—the simple ordering. The motivation to do this is to use the Kuhn-Tucker equivalence theorem to prove the famous pool-adjacent-violators algorithm (PAVA) and to give the graphical representation—greatest convex minorant (GCM).

Now we still consider the generalized isotonic regression problem. However, the specified partial ordering is the simple ordering. That is,

$$\begin{aligned} & \text{Minimize } \sum_{i=1}^k [\Phi(x_i) - g_i x_i] w_i, \\ & \text{subject to } x_1 \leq x_2 \leq \cdots \leq x_k. \end{aligned} \tag{6.1}$$

Note that $[\Phi(x) - gx]$ is a convex function of x . Let

$$\Psi(\mathbf{x}, \lambda) = - \sum_{i=1}^k [\Phi(x_i) - g_i x_i] w_i + \sum_{i=1}^{k-1} \lambda_i (x_{i+1} - x_i).$$

Using the Kuhn-Tucker equivalence theorem, the solution to problem (6.1) is \mathbf{x}^* if and only if there exists a vector λ such that

$$(6.2) \left\{ \begin{array}{l} \frac{\partial \Psi}{\partial x_i} \Big|_{\mathbf{x}^*} = [g_i - \varphi(x_i^*)] w_i + \lambda_{i-1} - \lambda_i = 0, \quad i = 1, \dots, k \\ \frac{\partial \Psi}{\partial \lambda_i} \Big|_{\mathbf{x}^*} = x_{i+1}^* - x_i^* \geq 0, \quad i = 1, \dots, k-1 \\ \lambda_i \left(\frac{\partial \Psi}{\partial \lambda_i} \Big|_{\mathbf{x}^*} \right) = \lambda_i (x_{i+1}^* - x_i^*) = 0, \quad i = 1, \dots, k-1 \\ \lambda_i \geq 0, \quad i = 1, \dots, k-1, \end{array} \right. \tag{6.2}^*$$

where we use the convention that $\lambda_0 = \lambda_k = 0$. Apparently, the Kuhn-Tucker conditions (6.2) are the special case of the Kuhn-Tucker conditions (4.3).

Graphical representation—greatest convex minorant

This very elegant algorithm is due to W. T. Reid (cf. Brunk (1956)). Plot the points $P_j = (W_j, G_j)$, $j = 0, 1, \dots, k$ with $W_j = \sum_{i=1}^j w_i$ and $G_j = \sum_{i=1}^j g_i w_i$ for $j = 1, 2, \dots, k$ and $P_0 = (0, 0)$. The plot of these points is called the *cumulative sum diagram* (CSD) for the given function g with the weights w . Note that the slope of the segment joining P_{j-1} to P_j is g_j , $j = 1, 2, \dots, k$. Let G^* be the *greatest convex minorant* (GCM) of the CSD over the interval $[0, W_k]$. The value, G_j^* , is the supremum of the values, at j , of all convex functions which lie entirely below the CSD. It is straightforward to see that the function defined in this way is convex over $[0, W_k]$. A graph of the CSD and GCM for a given set of points is given in Figure 6.1.

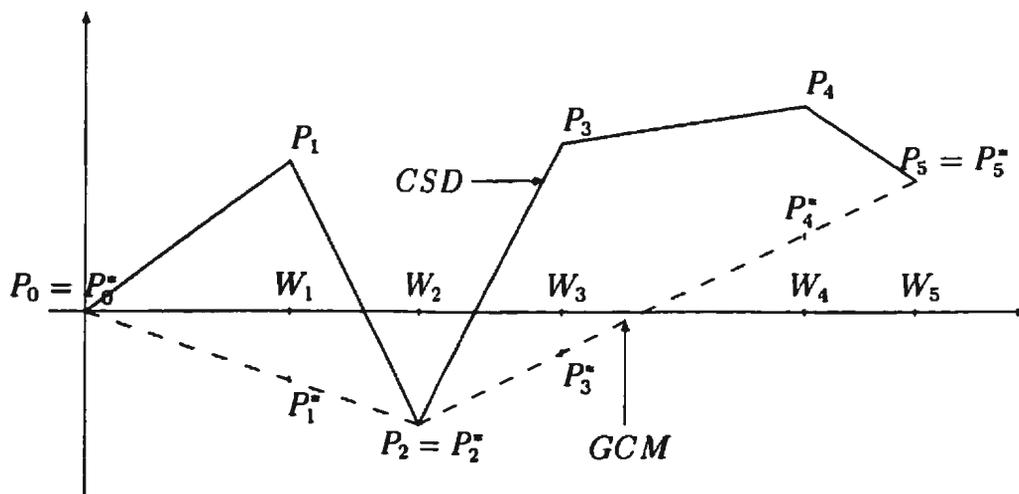


Figure 6.1. The cumulative sum diagram (CSD) and the greatest convex minorant (GCM)

Theorem 6.1. If \mathbf{x}^* is the solution to problem (6.1), then $(\varphi(x_1^*), \varphi(x_2^*), \dots, \varphi(x_k^*))$ is the left-hand slope of the greatest convex minorant.

Proof. For the simple ordering, the sets $\{1, \dots, j\}$, $j = 1, \dots, k$ are all lower sets. Among the equations (6.2)* we sum some of them together corresponding to all lower sets respectively. Then the Kuhn-Tucker conditions (6.2) imply

$$\left\{ \begin{array}{l} \sum_{i=1}^j [g_i - \varphi(x_i^*)]w_i - \lambda_j = 0, \quad j = 1, \dots, k-1 \\ \sum_{i=1}^k [g_i - \varphi(x_i^*)]w_i = 0 \\ \varphi(x_1^*) \leq \varphi(x_2^*) \leq \dots \leq \varphi(x_k^*) \\ \lambda_i(\varphi(x_{i+1}^*) - \varphi(x_i^*)) = 0, \quad i = 1, \dots, k-1 \\ \lambda_i \geq 0, \quad i = 1, \dots, k-1 \end{array} \right. \quad (6.3)$$

since $\varphi(x)$ is nondecreasing in x .

Plot the points $P_j^* = (W_j, G_j^*)$, $j = 0, 1, \dots, k$ with $W_j = \sum_{i=1}^j w_i$ and $G_j^* = \sum_{i=1}^j \varphi(x_i^*)w_i$ for $j = 1, 2, \dots, k$ and $P_0 = (0, 0)$. The plot of these points is called PLOT G^* . Thus, from Conditions (6.3), we have

$$\left\{ \begin{array}{l} G_j \geq G_j^*, \quad j = 1, \dots, k-1 \\ G_k = G_k^* \\ \varphi(x_1^*) \leq \varphi(x_2^*) \leq \dots \leq \varphi(x_k^*), \end{array} \right.$$

which means PLOT G^* is a convex minorant of the CSD over the interval $[0, W_k]$.

Now we discuss some properties of PLOT G^* , which will be employed to show that

PLOT G^* is the greatest one among all convex minorants of the CSD. Clearly,

$$(1). G_0 = G_0^*, \quad G_k = G_k^*,$$

$$(2). G_{j-1} > G_{j-1}^* \text{ implies } x_{j-1}^* = x_j^*.$$

In fact, $G_{j-1} > G_{j-1}^* \implies \lambda_{j-1} > 0 \implies x_{j-1}^* = x_j^*$ by Conditions (6.3).

Let M be the non-empty subset of $\{0, 1, 2, \dots, k\}$ consisting of $m \in M, G_m = G_m^*$. Consider $j \in \Omega - M$. Then we have $G_j > G_j^*$. Let α and β be consecutive indices in M such that $\alpha < j < \beta$. Property (2) implies that P_j^* lies on the line segment between P_α^* and P_β^* .

Let G' be any convex minorant of CSD over the interval $[0, W_k]$. Then $G'_j \leq G_j, j = 1, 2, \dots, k$. Therefore, $G'_m \leq G_m^*$ for each $m \in M$. The convexity of G' over $[0, W_k]$ implies that $G'_j \leq G_j^*, j = 1, 2, \dots, k$. So PLOT G^* is the greatest convex minorant and $(\varphi(x_1^*), \varphi(x_2^*), \dots, \varphi(x_k^*))$ is the left-hand slope of PLOT G^* . \square

The pool-adjacent-violators algorithm (PAVA)

The most widely used algorithm for computing the isotonic regression for a simple order is the *pool-adjacent-violator algorithm* (PAVA) first published by Ayer et al. in 1955. We will use the Kuhn-Tucker equivalence theorem to prove this algorithm.

Lemma 6.1. If $g_i \geq g_{i+1}$, then $x_i^* = x_{i+1}^*$.

Proof. Assume that $x_i^* < x_{i+1}^*$. By the Kuhn-Tucker conditions (6.2), we have $\lambda_i = 0$.

Now we consider the relevant equations in (6.2)*

$$\begin{cases} (g_i - \varphi(x_i^*))w_i + \lambda_{i-1} = 0, \\ (g_{i+1} - \varphi(x_{i+1}^*))w_{i+1} - \lambda_{i+1} = 0. \end{cases}$$

Since all λ_i 's are nonnegative, $g_i < \varphi(x_i^*)$ and $\varphi(x_{i+1}^*) < g_{i+1}$. It follows that $\varphi(x_{i+1}^*) < \varphi(x_i^*)$ which contradicts the assumption that $x_i^* < x_{i+1}^*$. Therefore, $x_i^* = x_{i+1}^*$. \square

Theorem 6.2. The final solution of PAVA to $\mathbf{g} = (g_1, \dots, g_k)$ with respect to the weights $\mathbf{w} = (w_1, \dots, w_k)$ is $(\varphi(x_1^*), \dots, \varphi(x_k^*))$.

Proof. We check the Kuhn-Tucker conditions (6.2). If $g_1 < \dots < g_k$, then all λ_i 's are zero. Hence, $(\varphi(x_1^*), \dots, \varphi(x_k^*)) = (g_1, \dots, g_k)$. Otherwise, there exists $g_i \geq g_{i+1}$, $1 \leq i \leq k-1$. By Lemma 6.1, we know that $x_i^* = x_{i+1}^*$. Then problem (6.1) is reduced to

$$\text{Minimize} \quad \sum_{\substack{j=1 \\ j \neq i, i+1}}^k [\Phi(x_j) - g_j x_j] w_j + [\Phi(x) - g_i x] w_i + [\Phi(x) - g_{i+1} x] w_{i+1},$$

$$\text{subject to } x_1 \leq \dots \leq x_{i-1} \leq x \leq x_{i+2} \leq \dots \leq x_k.$$

Also,

$$[\Phi(x) - g_i x] w_i + [\Phi(x) - g_{i+1} x] w_{i+1} = [\Phi(x) - \bar{g}_{i,i+1} x] (w_i + w_{i+1}),$$

where $\bar{g}_{i,i+1} = (w_i g_i + w_{i+1} g_{i+1}) / (w_i + w_{i+1})$.

Therefore, problem (6.1) is equivalent to

$$\text{Minimize} \quad \sum_{\substack{j=1 \\ j \neq i, i+1}}^k [\Phi(x_j) - g_j x_j] w_j + [\Phi(x) - \bar{g}_{i,i+1} x] (w_i + w_{i+1}), \quad (6.4)$$

$$\text{subject to } x_1 \leq \dots \leq x_{i-1} \leq x \leq x_{i+2} \leq \dots \leq x_k.$$

Using Lemma 6.1, we repeat the above process. This is just the PAVA. And the solution of PAVA is $(\varphi(x_1^*), \dots, \varphi(x_k^*))$. \square

Two other algorithms, namely the *minimum-lower-sets algorithm* and the *min-max formulas*, have been used extensively in deriving properties of the isotonic regression. As with the PAVA, they are both straightforward consequences of the graphical representation of the isotonic regression for the simply ordered case. However, these algorithms hold in the more general context discussed in the last section.

7. MULTINOMIAL ESTIMATION UNDER STOCHASTIC ORDERING

Consider an experiment, the outcome of which must be one of k mutually exclusive events which have probabilities p_1, \dots, p_k and suppose these probabilities are to be estimated. If n independent trials of this experiment are performed, then p_i could be estimated by the relative frequency \hat{p}_i of the event which has probability p_i . However, in some situations, one may believe that the p_i 's satisfy certain order restrictions and may desire estimates which also satisfy these restrictions. For instance, C. G. Park, through a private communication, considers the problem that the p_i 's satisfy

$$p_1 - q_1 \leq \dots \leq p_k - q_k, \quad (7.1)$$

where $\mathbf{q} = (q_1, \dots, q_k)$ is a given probability vector. We will discuss the maximum likelihood estimate and Neyman modified minimum chi-square estimate of \mathbf{p} subject to some constraints. This will give us some insights into the applications of the Kuhn-Tucker equivalence theorem.

7.1. Maximum Likelihood Estimate

We will discuss the special case of (7.1) in $k = 3$ here. Suppose x_1, x_2, x_3 are observed random variables which possess a multinomial distribution with parameter n and probability vector (PV) \mathbf{p} . Assume also that \mathbf{p} satisfies the restriction (7.1). We wish to estimate \mathbf{p} .

The maximum likelihood estimate problem can be expressed by the problem

$$\text{Maximize } \sum_{i=1}^3 x_i \ln p_i, \quad (7.2)$$

$$\text{subject to } p_1 - q_1 \leq p_2 - q_2 \leq p_3 - q_3.$$

Since the function $\ln p$ is strictly concave and $K = \{\mathbf{p} : p_1 - q_1 \leq p_2 - q_2 \leq p_3 - q_3, p_1 + p_2 + p_3 = 1\}$ is a closed convex cone, the solution to problem (7.2) exists uniquely. Let

$$\begin{aligned} \Psi(\mathbf{p}, \lambda) = & x_1 \ln p_1 + x_2 \ln p_2 + x_3 \ln p_3 + \lambda_1 [(p_2 - q_2) - (p_1 - q_1)] \\ & + \lambda_2 [(p_3 - q_3) - (p_2 - q_2)] + \lambda_3 (p_1 + p_2 + p_3 - 1). \end{aligned}$$

By the Kuhn-Tucker equivalence theorem, \mathbf{p}^* is the solution to problem (7.2) if and only if there exists a vector $\lambda = (\lambda_1, \lambda_2, \lambda_3)$ such that

$$\left\{ \begin{array}{l} \frac{x_1}{p_1^*} - \lambda_1 + \lambda_3 = 0 \\ \frac{x_2}{p_2^*} - \lambda_2 + \lambda_1 + \lambda_3 = 0 \\ \frac{x_3}{p_3^*} + \lambda_2 + \lambda_3 = 0 \\ \lambda_1 [(p_2^* - q_2) - (p_1^* - q_1)] = 0 \\ \lambda_2 [(p_3^* - q_3) - (p_2^* - q_2)] = 0 \\ p_1^* - q_1 \leq p_2^* - q_2 \leq p_3^* - q_3 \\ \lambda_1 \geq 0, \quad \lambda_2 \geq 0 \\ p_1^* + p_2^* + p_3^* = 1. \end{array} \right. \quad (7.3)$$

From the constraint $p_1 - q_1 \leq p_2 - q_2 \leq p_3 - q_3$, we may divide the solution \mathbf{p}^* into four cases. We will discuss these four cases of the solution respectively as follows.

Case I. $p_1^* - q_1 < p_2^* - q_2 < p_3^* - q_3$.

From the Kuhn-Tucker conditions (7.3), we have $\lambda_1 = \lambda_2 = 0$. So

$$\begin{aligned} \frac{x_1}{p_1^*} &= \frac{x_2}{p_2^*} = \frac{x_3}{p_3^*} \\ \Rightarrow \begin{cases} p_2^* = \frac{x_2 p_1^*}{x_1} \\ p_3^* = \frac{x_3 p_1^*}{x_1} \end{cases} \\ \Rightarrow p_1^* + \frac{x_2 p_1^*}{x_1} + \frac{x_3 p_1^*}{x_1} &= 1 \\ \Rightarrow p_1^* = \hat{p}_1, \quad p_2^* = \hat{p}_2, \quad p_3^* = \hat{p}_3. \end{aligned}$$

Substituting $\mathbf{p}^* = \hat{\mathbf{p}}$ into the Kuhn-Tucker conditions (7.3), we conclude that

$$\mathbf{p}^* \text{ has the format of Case I} \iff \hat{p}_1 - q_1 < \hat{p}_2 - q_2 < \hat{p}_3 - q_3. \quad (7.4)$$

Moreover, $\mathbf{p}^* = \hat{\mathbf{p}}$.

Case II. $p_1^* - q_1 = p_2^* - q_2 = p_3^* - q_3$.

On the basis of the Kuhn-Tucker conditions (7.3), we obtain $\lambda_1 \geq 0$, $\lambda_2 \geq 0$. In this case,

$$\begin{aligned} p_1^* - q_1 &= p_2^* - q_2 = p_3^* - q_3 \\ \Rightarrow \begin{cases} p_2^* = p_1^* - q_1 + q_2 \\ p_3^* = p_1^* - q_1 + q_3 \end{cases} \\ \Rightarrow p_1^* + (p_1^* - q_1 + q_2) + (p_1^* - q_1 + q_3) &= 1 \\ \Rightarrow p_1^* = q_1, \quad p_2^* = q_2, \quad p_3^* = q_3. \end{aligned}$$

Substitute $\mathbf{p}^* = \mathbf{q}$ into the Kuhn-Tucker conditions (7.3) which become

$$\begin{cases} \frac{x_1}{q_1} - \lambda_1 + \lambda_3 = 0 \\ \frac{x_2}{q_2} - \lambda_2 + \lambda_1 + \lambda_3 = 0 \\ \frac{x_3}{q_3} + \lambda_2 + \lambda_3 = 0 \\ \lambda_1 \geq 0, \quad \lambda_2 \geq 0. \end{cases} \quad (7.5)$$

Summing the first three equations in Conditions (7.5), we solve

$$\lambda_3 = -\frac{1}{3}\left(\frac{x_1}{q_1} + \frac{x_2}{q_2} + \frac{x_3}{q_3}\right).$$

Replace λ_3 by the above expression in Conditions (7.5) and sum the first two equations. Then Conditions (7.5) are equivalent to

$$\begin{cases} \frac{1}{3}\left(\frac{2x_1}{q_1} - \frac{x_2}{q_2} - \frac{x_3}{q_3}\right) - \lambda_1 = 0 \\ \frac{1}{3}\left(\frac{x_1}{q_1} + \frac{x_2}{q_2} - \frac{2x_3}{q_3}\right) - \lambda_2 = 0 \\ \lambda_1 \geq 0, \quad \lambda_2 \geq 0, \end{cases}$$

which are equivalent to

$$\begin{cases} \frac{2x_1}{q_1} \geq \frac{x_2}{q_2} + \frac{x_3}{q_3} \\ \frac{x_1}{q_1} + \frac{x_2}{q_2} \geq \frac{2x_3}{q_3}. \end{cases} \quad (7.6)$$

Case III. $p_1^* - q_1 = p_2^* - q_2 < p_3^* - q_3$.

That means $\lambda_2 = 0$, $\lambda_1 \geq 0$. So the Kuhn-Tucker conditions (7.3) are changed into

$$\left\{ \begin{array}{l} \frac{x_1}{p_1^*} - \lambda_1 + \lambda_3 = 0 \\ \frac{x_2}{p_2^*} + \lambda_1 + \lambda_3 = 0 \\ \frac{x_3}{p_3^*} + \lambda_3 = 0 \\ p_1^* + p_2^* + p_3^* = 1 \\ p_1^* - q_1 = p_2^* - q_2 < p_3^* - q_3 \\ \lambda_1 \geq 0. \end{array} \right. \quad (7.7)$$

Solving Conditions (7.7), we obtain a quadratic equation for p_1^* as

$$2np_1^{*2} + [(3x_1 + x_2 + 2x_3)(q_2 - q_1) - (x_1 + x_2)]p_1^* + x_1(q_2 - q_1)[(q_2 - q_1) - 1] = 0.$$

Note that the solution \mathbf{p}^* to problem (7.2) exists uniquely. Solving the above equation, we can get two candidates of solution p_1^* , and then obtain two corresponding candidates of p_2^* and p_3^* by

$$p_2^* = p_1^* - q_1 + q_2,$$

$$p_3^* = 1 - 2p_1^* + q_1 - q_2.$$

It is not difficult to delete one unreasonable candidate of the solution by checking Conditions (7.7). Thus we finally obtain the unique solution to problem (7.2).

Case IV. $p_1^* - q_1 < p_2^* - q_2 = p_3^* - q_3$.

This is the situation analogous to *Case III*. Here $\lambda_1 = 0$, $\lambda_2 \geq 0$. We can write the

Kuhn-Tucker conditions (7.3) as

$$\left\{ \begin{array}{l} \frac{x_1}{p_1^*} + \lambda_3 = 0 \\ \frac{x_2}{p_2^*} - \lambda_2 + \lambda_3 = 0 \\ \frac{x_3}{p_3^*} + \lambda_2 + \lambda_3 = 0 \\ p_1^* + p_2^* + p_3^* = 1 \\ p_1^* - q_1 < p_2^* - q_2 = p_3^* - q_3 \\ \lambda_2 \geq 0. \end{array} \right. \quad (7.8)$$

Solving Conditions (7.8), we can obtain a quadratic equation for p_2^* as

$$2np_2^{*2} + [(2x_1 + 3x_2 + x_3)(q_3 - q_2) - (x_2 + x_3)]p_2^* + x_2(q_3 - q_2)[(q_3 - q_2) - 1] = 0.$$

A similar discussion to *Case III* makes us get a unique solution \mathbf{p}^* to problem (7.2).

Remark 7.1.

1. Although we do not provide the exact expression of solution \mathbf{p}^* to problem (7.2) in *Case III* and *IV*, it is not hard to solve those quadratic equations and get the unique solution to problem (7.2).

2. We proposed the two regions (7.4) and (7.6) corresponding to the solutions of *Case I* and *Case II*, respectively. These two regions also divide the sample space into four regions. The other two regions correspond to the solutions of *Case III* and *Case IV*, respectively. Figure 7.1 illustrates the relationship between sample regions and formats of solution \mathbf{p}^* .

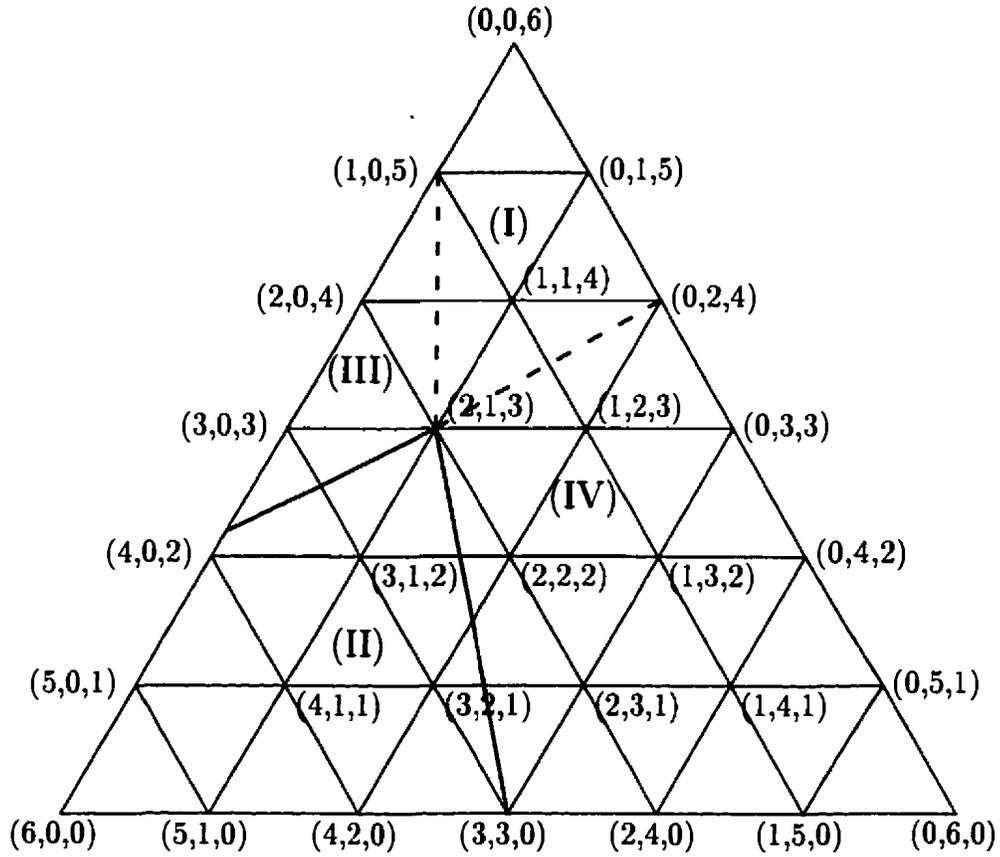


Figure 7.1. The relationship between sample regions and formats of solution \mathbf{p}^*

Here $n = 6$.

$$\mathbf{q} = (2/6, 1/6, 3/6).$$

(I). $p_1^* - q_1 < p_2^* - q_2 < p_3^* - q_3$ corresponding to Region (7.4): $x_1 - 2 < x_2 - 1 < x_3 - 3$.

(II). $p_1^* - q_1 = p_2^* - q_2 = p_3^* - q_3$ corresponding to Region (7.6):
$$\begin{cases} 3x_1 \geq 3x_2 + x_3 \\ 3x_1 + 6x_2 \geq 4x_3. \end{cases}$$

(III). $p_1^* - q_1 = p_2^* - q_2 < p_3^* - q_3$.

(IV). $p_1^* - q_1 < p_2^* - q_2 = p_3^* - q_3$.

Example 7.1. We will now illustrate how to solve the solution \mathbf{p}^* when \mathbf{p}^* has the format of *Case III* or *Case IV*. Given $\mathbf{q} = (2/6, 1/6, 3/6)$, we have the data

$$\mathbf{x} = (1, 4, 1), \quad n = 6$$

$$\hat{\mathbf{p}} = (1/6, 4/6, 1/6).$$

From Figure 7.1, we can conclude that the solution \mathbf{p}^* has the format of *Case IV*.

So we obtain the quadratic equation for p_2^* as

$$27p_2^{*2} - 2 = 0$$

Solving the above equation, we can get two candidates of p_2^* as follows

$$p_2^* = .272 \quad \text{or} \quad p_2^* = -.272.$$

Since $p_2^* = -.272 < 0$, this candidate is unreasonable. Hence, $p_2^* = .272$. Then

$$p_3^* = p_2^* + (q_3 - q_2) = .605,$$

$$p_1^* = 1 - p_2^* - p_3^* = .123.$$

Checking the Kuhn-Tucker conditions (7.3), we obtain the solution to problem (7.2) as

$$\mathbf{p}^* = (.123, .272, .605).$$

7.2. Neyman Modified Minimum Chi-square Estimate

The Neyman modified minimum chi-square estimation technique for \mathbf{p} leads to the optimization problem which can be phrased, in the following manner:

$$\text{minimize} \quad \sum_{i=1}^k \frac{(x_i - np_i)^2}{x_i}.$$

Also,

$$\sum_{i=1}^k \frac{(x_i - np_i)^2}{x_i} = n \sum_{i=1}^k \frac{(\hat{p}_i - p_i)^2}{\hat{p}_i}.$$

Now we consider the above optimization problem. We add the constraint that $\mathbf{p} - \mathbf{q}$ satisfies a specified partial ordering. Thus, this optimization problem is equivalent to

$$\text{Maximize } -\frac{1}{2} \sum_{i=1}^k \frac{(\hat{p}_i - p_i)^2}{\hat{p}_i}, \quad (7.9)$$

$$\text{subject to } p_i - q_i \leq p_j - q_j \text{ when } i \prec j,$$

where both \mathbf{p} and \mathbf{q} (given) are probability vectors, and the binary relation \prec is a specified partial ordering. We can write the objective function in (7.9) as

$$-\frac{1}{2} \sum_{i=1}^k \frac{[(\hat{p}_i - q_i) - (p_i - q_i)]^2}{\hat{p}_i}.$$

Let $\mathbf{r} = \mathbf{p} - \mathbf{q}$ and $\hat{\mathbf{r}} = \hat{\mathbf{p}} - \mathbf{q}$. Then problem (7.9) is equivalent to

$$\text{Maximize } -\frac{1}{2} \sum_{i=1}^k \frac{(\hat{r}_i - r_i)^2}{\hat{p}_i},$$

$$\text{subject to } r_i \leq r_j \text{ when } i \prec j, \quad (7.10)$$

$$\text{and } \sum_{i=1}^k r_i = 0.$$

Note that $-\frac{(\hat{r}_i - r_i)^2}{2\hat{p}_i}$ is a strictly concave function of r_i . Using the notations for partial orderings, let

$$\Psi(\mathbf{r}, \lambda) = -\frac{1}{2} \sum_{i=1}^k \frac{(\hat{r}_i - r_i)^2}{\hat{p}_i} + \sum_{(i,j) \in E} \lambda_{ij}(r_j - r_i) + \lambda_0 \left(\sum_{i=1}^k r_i \right).$$

Using the Kuhn-Tucker equivalence theorem, we conclude that \mathbf{r}^* is the solution

to problem (7.10) if and only if there exists a vector λ such that

$$(7.11) \left\{ \begin{array}{l} (\hat{r}_i + \lambda_0 \hat{p}_i - r_i^*) \cdot \frac{1}{\hat{p}_i} + \sum_{j \in P_i} \lambda_{ji} - \sum_{j \in S_i} \lambda_{ij} = 0, \quad i = 1, \dots, k \\ \lambda_{ij}(r_j^* - r_i^*) = 0, \quad (i, j) \in E \\ r_i^* \leq r_j^*, \quad (i, j) \in E \\ \lambda_{ij} \geq 0, \quad (i, j) \in E \\ \sum_{i=1}^k r_i^* = 0, \end{array} \right\} (7.11)^*$$

where

$$S_i = \{j : (i, j) \in E\},$$

$$P_i = \{j : (j, i) \in E\}, \quad i \in \Omega.$$

Similarly to the proof of Theorem 4.1, we have $\mathbf{r}^* = E_{\hat{\mathbf{p}}^{-1}}(\hat{\mathbf{r}} + \lambda_0 \hat{\mathbf{p}}|K)$ from Conditions (7.11)*, where $K = \{\mathbf{x} : x_i \leq x_j, (i, j) \in E\}$. Thus, the solution \mathbf{r}^* to problem (7.10) is given by $E_{\hat{\mathbf{p}}^{-1}}(\hat{\mathbf{r}} + \lambda_0 \hat{\mathbf{p}}|K)$ provided that

$$\sum_{i=1}^k E_{\hat{\mathbf{p}}^{-1}}(\hat{\mathbf{r}} + \lambda_0 \hat{\mathbf{p}}|K) = 0.$$

The key to obtaining the solution \mathbf{r}^* is to determine λ_0 . Now we discuss the property of λ_0 so that we can get the solution \mathbf{r}^* more effectively. Note that $\hat{\mathbf{r}} = \hat{\mathbf{p}} - \mathbf{q}$ and that $\hat{\mathbf{r}} + \lambda_0 \hat{\mathbf{p}} = (1 + \lambda_0)\hat{\mathbf{p}} - \mathbf{q}$ is increasing in λ_0 . Since $\min_i \{\hat{r}_i + \lambda_0 \hat{p}_i\} \leq r_i^* \leq \max_i \{\hat{r}_i + \lambda_0 \hat{p}_i\}$ by Theorem 2.2, there exist λ_0^1 and λ_0^2 such that $\min_i \{\hat{r}_i + \lambda_0^1 \hat{p}_i\} \geq 0$ and $\max_i \{\hat{r}_i + \lambda_0^2 \hat{p}_i\} \leq 0$. Under the restriction $\sum_{i=1}^k r_i^* = 0$, we limit $\lambda_0 \in [\lambda_0^1, \lambda_0^2]$.

Since much is known about computing the isotonic regression (cf. Robertson et al. (1988)), the Neyman modified minimum chi-square estimate of \mathbf{p} is tractable when

dealing with isotonic cones.

Example 7.2. Consider $k = 3$ and the specified partial ordering is the simple ordering. We will determine λ_0 and calculate the approximate solution \mathbf{r}^* and \mathbf{p}^* .

We have the data and \mathbf{q} as follows.

$$\mathbf{q} = (2/6, 1/6, 3/6)$$

$$\mathbf{x} = (1, 4, 1), \quad n = 6$$

$$\hat{\mathbf{p}} = (1/6, 4/6, 1/6)$$

$$\hat{\mathbf{p}}^{-1} = (6, 1.5, 6)$$

$$\hat{\mathbf{r}} = \hat{\mathbf{p}} - \mathbf{q} = (-1/6, 3/6, -2/6).$$

First of all, we determine the interval $[\lambda_0^1, \lambda_0^2]$ of λ_0 . By the data, we have

$$\hat{\mathbf{r}} + \lambda_0 \hat{\mathbf{p}} = \left(\frac{\lambda_0 - 1}{6}, \frac{4\lambda_0 + 3}{6}, \frac{\lambda_0 - 2}{6} \right).$$

Then $[\lambda_0^1, \lambda_0^2] = [-.75, 2]$. Table 7.1 presents the choice of λ_0 using linear interpolation and the approximate solution \mathbf{r}^* .

TABLE 7.1. THE CALCULATION OF APPROXIMATE SOLUTION \mathbf{r}^*

λ_0	$\hat{\mathbf{r}} + \lambda_0 \hat{\mathbf{p}}$	\mathbf{r}^*	$\sum_{i=1}^3 r_i^*$
-.7500	(-.2917, .0000, -.4583)	(-.3333, -.3333, -.3333)	-.9999
2.0000	(.1667, 1.8333, .0000)	(.1667, .3667, .3667)	.9001
.6972	(-.0505, .9648, -.2171)	(-.0505, .0193, .0193)	-.0119
.7141	(-.0477, .9761, -.2143)	(-.0477, .0238, .0238)	-.0001
.7142	(-.0476, .9761, -.2143)	(-.0476, .0238, .0238)	.0000

From Table 7.1, we take $\lambda_0 = .7142$ and obtain the approximate solution $\mathbf{r}^* = (-.0476, .0238, .0238)$. Thus, the approximate solution $\mathbf{p}^* = \mathbf{r}^* + \mathbf{q} = (.2857, .1905, .5238)$ is the Neyman modified minimum chi-square estimate of \mathbf{p} .

Remark 7.2. In Section 5, we mentioned the divergence of the PV $\hat{\mathbf{p}}$ with respect to the PV \mathbf{p} of order λ defined by Cressie and Read (1984) as

$$I^\lambda(\hat{\mathbf{p}} : \mathbf{p}) = \frac{1}{\lambda(\lambda + 1)} \sum_{i=1}^k \hat{p}_i \left[\left(\frac{\hat{p}_i}{p_i} \right)^\lambda - 1 \right]. \quad (7.13)$$

In particular, the maximum likelihood estimate and the Neyman modified minimum chi-square estimate of \mathbf{p} correspond to the two special cases of the optimization problem $\min 2nI^\lambda(\hat{\mathbf{p}} : \mathbf{p})$ when $\lambda = 0$ and $\lambda = -2$, respectively. As discussed in Dykstra and Lee (1991), if $\lambda = 1$, the solution to $\min 2nI^\lambda(\hat{\mathbf{p}} : \mathbf{p})$ is the Pearson minimum chi-square estimate of \mathbf{p} . And if $\lambda = -1$, $\min 2nI^\lambda(\hat{\mathbf{p}} : \mathbf{p})$ means the estimation technique of minimum discriminant information by continuity in λ . Unfortunately, when we add the simply ordered constraint on $\mathbf{p} - \mathbf{q}$, it is very difficult to use the Kuhn-Tucker equivalence theorem to solve the latter two optimization problems even when $k = 3$. Apparently, there is no one theorem to solve every problem. This encourages us to continuously pursue other powerful methods to solve our problems.

So far we have provided some insights into the applications of the Kuhn-Tucker equivalence theorem. This equivalence theorem itself solves inequality constrained maximization problems, and we believe it will be extensively applied as a powerful and effective tool to order restricted statistical inference. Clearly, the more tools that are available for solving a problem, the better the possibility of attaining success. The Kuhn-Tucker equivalence theorem gives a series of equations and inequalities for the solution to order restricted problems. At this point, it is very intuitive to use

this equivalence theorem to solve these problems. Moreover, its application has made many proofs much simpler compared to the other proof approaches.

References

- Ayer, M., H. D. Brunk, G. M. Ewing, W. T. Reid, and E. Silverman (1955). An empirical distribution function for sampling with incomplete information, *Ann. Math. Statist.*, **26**, 641-7.
- Barlow, R. E., D. J. Bartholomew, J. M. Bremner and H. D. Brunk (1972). *Statistical Inference under Order Restrictions*, Wiley, New York.
- Barlow, R. E. and H. D. Brunk (1972). The isotonic regression problem and its dual, *J. Amer. Statist. Assoc.*, **67**, 140-147.
- Bartholomew, D. J. (1959). A test of homogeneity for ordered alternatives I and II, *Biometrika*, **46**, Nos. 1 and 2, 36-48, 328-35.
- Bickel, J. P. and K. A. Doksum (1977). *Mathematical Statistics*, Holden-Day, Oakland.
- Brunk, H. D. (1955). Maximum likelihood estimates of monotone parameters, *Ann. Math. Statist.*, **26**, 607-16.
- Brunk, H. D. (1956). On an inequality for convex functions, *Proc. Amer. Math. Soc.*, **7**, 817-24.
- Brunk, H. D. (1958). On the estimation of parameters restricted by inequalities, *Ann. Math. Statist.*, **29**, 437-54.

- Brunk, H. D. (1965). Conditional expectation given a σ -lattice and applications, *Ann. Math. Statist.*, **36**, 1339-50.
- Brunk, H. D., G. M. Ewing, and W. R. Utz (1957). Minimizing integrals in certain classes of monotone functions, *Pacific J. Math.*, **7**, 833-47.
- Cressie, N. and T. R. C. Read (1984). Multinomial goodness of fit models, *J. Roy Statist. Soc. (B)*, **46**, 440-464.
- Dykstra, R. L. (1983). An algorithm for restricted least squares regression, *J. Amer. Statist. Assoc.*, **78**, 837-42.
- Dykstra, R. L. (1985). Computational aspects of I-projections, *J. Statist. Comput. Simul.*, **21**, 265-274.
- Dykstra, R. L. and C. I. C. Lee (1991). Multinomial estimation procedures for isotonic cones. *Statist. & Probab. Lett.*, **11**, 155-160.
- Dykstra, R. L. and T. Robertson (1982). An algorithm for isotonic regression for two or more independent variables, *Ann. Statist.*, **10**, 708-11.
- Eeden, C. Van (1956). Maximum likelihood estimation of ordered probabilities, *Indagationes Mathematicae*, **18**, No. 3, 444-455.
- Eeden, C. Van (1957a). Maximum likelihood estimation of partially or completely ordered parameters, I. *Indagationes Mathematicae*, **19**, No. 2, 128-36.

- Eeden, C. Van (1957b). Maximum likelihood estimation of partially or completely ordered parameters, II. *Indagationes Mathematicae*, **19**, No. 2, 201-11.
- Eeden, C. Van (1957c). Note on two methods for estimating ordered parameters of probability distributions, *Indagationes Mathematicae*, **19**, No. 3, 506-12.
- Eeden, C. Van (1958). Testing and estimation ordered parameters of probability distributions, (*Doctoral Dissertation, University of Amsterdam*), Studentendrukkerij Poortpers, Amsterdam.
- Farebrother, R. W. (1986). Testing linear inequality constraints in the standard linear model, *Comm. Statist. (A)*, **15**, 7-31.
- Good, I. J. (1963). Maximum entropy for hypothesis formulation, especially for multidimensional contingency tables, *Ann. Math. Statist.*, **34**, 911-34.
- Gourieroux, C., A. Holly, and A. Monfort (1982). Likelihood ratio test, Wald test and Kuhn-Tucker test in linear model with inequality constraints on the regression parameters, *Econometrica*, **50**, 63-79.
- Jaynes, E. T. (1957). Information theory and statistical mechanics, *Phys. Rev.*, **106**, 620-30.
- Kuhn, H. W. and A. W. Tucker (1951), Nonlinear programming, 2nd Berkeley Symp., *Math. Statist. probab.* (University of California Press, Berkeley) 481-492.

- Kullback, S. (1959). *Information Theory and Statistics*, Wiley, New York.
- Lee, C. I. C. (1983). The min-max algorithm and isotonic regression, *Ann. Statist.*, **11**, 467-77.
- Rao, C. R. (1965). *Linear Statistical Inference and its Applications*, Wiley, New York.
- Robertson, T., F. T. Wright and R. L. Dykstra (1988). *Order Restricted Statistical Inference*, Wiley, Chichester.
- Rockafellar, R. T. (1970). *Convex Analysis*, Princeton University Press.
- Sanov, I. N. (1957). On the probability of large deviations of random variables, *Mat. Sb.*, **42**, 11-44.
- Wolak, F. A. (1987). An exact test for multiple inequality and equality constraints in the linear regression model, *J. Amer. Statist. Assoc.*, **82**, 782-93.





